Hiding Lemons among Peaches: Optimal Retention and Promotion Policy Design

Andrew B. Choi and Takuma Habu*

20 March 2025

– Click here for latest version –

Abstract

How should an employer design its retention and promotion policy when there is competition for its employees? In an environment where the incumbent employer learns more about its workers than other potential employers, we show that the incumbent's optimal policy always over-retains workers—inefficiently retaining workers who would be more productive elsewhere—to dampen the positive signalling effect of retention. In contrast, the optimal policy may over- or under-promote workers. The incumbent's incentive to distort its policy is driven by the technologies of competing firms in the labour market. We also demonstrate the extent to which the firm can implement the optimal policy without the ability to commit and study the incentive for the firm to manipulate the signalling effect via designing jobs. Our results shed light on the role of (possibly vacuous) job titles and provide a novel rationale for the Peter Principle.

*Andrew B. Choi: Department of Economics, University of Michigan, andybchoi@gmail.com. Takuma Habu: Department of Economics, Cornell University, takumahabu@cornell.edu. We thank Marco Battaglini, Raphael Boleslavsky, Bart de Koning, Yi Chen, Thomas Jungbauer, Emir Kamenica, Michael Gibbs, Robert Gibbons, Justin Johnson, Rohit Lamba, Xiao Lin, Elliot Lipnowski, Doron Ravid, Collin Raymond, Lars Stole, Michael Waldman, and Kai Hao Yang for helpful comments.

How should a firm retain or promote workers who can leave for another employer? When an employer has private information about its workers, retention or promotion decisions can drive up the wage because they serve as positive signals about workers' productivity to other potential employers.[1] In this paper, we explore whether a firm would dampen the positive signalling effect by designing a policy that retains or promotes some "lemon" workers along with some "peach" workers despite the costs of doing so. After all, workplaces abound with stories of incompetent managers.

We show that the firm always finds it optimal to put in place a policy that over-retains workers; that is, the firm retains workers who would be more productive if employed elsewhere. However, whether the firm's optimal policy over- or under-promotes workers depends on how other potential employers value workers. The results differ for retention and promotion policies because promoting a worker to a more senior position entails an opportunity cost—the firm loses a worker in the worker's previous position—that is not present with retention.

To our knowledge, our paper is the first to study the signalling effect of retention and promotion jointly.[2] Moreover, our main result on promotion stands in contrast to the existing literature beginning with Waldman (1984) that finds that the signalling effect of promotion always leads to under-promotions (Waldman, 1984; Milgrom and Oster, 1987; Gibbons and Katz, 1991; Bernhardt, 1995; Ishida, 2004; Owan, 2004; Gibbons and Waldman, 2006; DeVaro and Waldman, 2012). The difference arises for two main reasons. First, we consider the firm's ex ante incentive to commit to promotion policies, whereas existing models have studied the firm's ex post optimal promotion decisions.[3] Second, we introduce heterogeneity in the production technologies of other potential employers. We therefore interpret our result as relevant for a firm wishing to design a retention and promotion policy before it learns the productivity of its workers, as is often the case in reality.

[1] Empirically, the literature has found evidence for the presence of signalling effect of promotions (DeVaro and Waldman, 2012; Bognanno and Melero, 2016; Cassidy, DeVaro and Kauhanen, 2016; Dato et al., 2016). Others also find evidence for asymmetric learning in labour markets (e.g., Gibbons and Katz, 1991; Schonberg, 2007; Pinkston, 2009; Kahn, 2013).

[2] The labour literature has long recognised the inefficiency arising from the signalling effect of promotions starting with Waldman (1984)—indeed, the term "promotion as signals" has become widely used (Gibbons and Roberts, 2013). With respect to retention, beginning with Greenwald (1986), the literature has predominantly studied the adverse selection effect arising from the employer's strategic decision to let go of "worse" workers when responding to outside offers. Gibbons and Katz (1991) is an exception. While their model captures both the signalling and adverse selection effects of retention, they do not consider promotions.

[3] An exception is Ghosh and Waldman (2010) who consider a model in which the firm can (only) commit to letting go of workers.

Our results also provide a novel rationale for the popular idea that people in a hierarchy tend to rise to "a level of respective incompetence," commonly referred to as the *Peter Principle* (Peter and Hull, 1969).[4] Our results suggest that, in cases where the labour market cannot distinguish incompetent workers from competent ones, it can be profitable for the firm to tolerate the presence of incompetent workers as a means to minimise wage costs.[5]

In our model, an incumbent firm employs a unit mass of junior workers with heterogeneous abilities. The firm must decide who to retain and, among those retained, who to promote to a senior position. The output from a worker in a given position increases linearly in the worker's ability. Relative to the junior position, the senior position's output increases more steeply in the worker's ability. A worker produces more in the senior position than in the junior position if and only if the worker's ability is above a threshold. The workers only care about the wage.[6]

The labour market consists of heterogeneous outside firms with linear production technologies. To retain a worker (either in a junior or a senior position), the incumbent firm pays a wage equal to the worker's outside option—the highest wage that the worker can command from outside firms. The outside firms only observe whether each worker was retained and, if so, whether the worker was promoted.[7]

The incumbent firm's problem is to maximize its profits by designing an *HR policy* that specifies retention and promotion probabilities as functions of a worker's ability. We are interested in how the firm's optimal policy differs from the efficient benchmark that would be obtained under complete information, where each worker is employed by a firm where they would be most productive.

Our first main finding (Theorem 1) is that, relative to the efficient benchmark, any optimal HR policy over-retains workers, and no optimal HR policy under-retains workers. That is, the firm always retains some workers who would be more productive at another firm and never lets go of workers who are most productive at the incumbent firm. Retaining a marginal worker is beneficial for the firm because it lowers the average ability, and thus the

[4] Benson, Li and Shue (2019) finds empirical support for the Peter Principle.

[5] Our interpretation of the principle is consistent with Lazear (2004). In a matching model with a secondary market, Vohra (2024) demonstrates that a firm may hire less qualified workers to deter poaching.

[6] As Baker, Jensen and Murphy (1988) notes, workers may care about retention and promotions for reasons other than wage. We assume away this aspect as in Waldman (1984).

[7] As in Waldman (1984), we do not introduce adverse selection issues related to how the incumbent firm responds to a wage offer from other firms (Greenwald, 1986; Gibbons and Katz, 1991) or allow the workers to possess private information (Costa, 1988).

wage cost, of all retained workers; however, doing so is also costly because the worker's output would be lower than the wage, which is based on the average ability of all retained workers. We show that the benefit strictly dominates the cost for the worker for whom it is just efficient for the firm to retain, which implies that the firm should retain workers with lower abilities. The result is driven by the heterogeneity of outside firms, which implies that the value of workers' outside options is not only increasing but also convex in ability.

Our second main finding (Theorem 2) is that an optimal HR policy may exhibit either over-promotion or under-promotion (or even both if the optimal policy is non-deterministic). The same rationale for over-retention means that over-promoting always increases the firm's profits from the promoted workers. However, unlike over-retention, over-promotion involves an opportunity cost—the marginally promoted worker, who had the highest ability among all junior workers, no longer works in the junior position. We show that the comparison between these two opposing effects depends on the nature of the heterogeneity among outside firms.

As an alternative way for the firm to attenuate the signalling effect of retention (with and without promotions), we consider the firm's incentive to design jobs by merging the junior and senior positions. We show that an employer cannot strictly benefit from merging positions, meaning the firm prefers its workers to specialise. Finally, we also discuss the incentive of the firm to create vacuous job titles to provide more information to the labour market.

Our model can be viewed as a signalling game where the sender commits to a signalling strategy. Our approach reveals that such signalling games can often be solved by establishing an equivalence with a related information design game in which the sender chooses a sequentially rational action after learning about the state from a public Blackwell experiment. This allows one to characterise optimal signalling strategies using tools from information design (Dworczak and Martini, 2019; Arieli et al., 2023). We emphasise that our solution approach applies more generally to settings beyond what we study in this paper.

The equivalence between the two games also means that the firm can indirectly implement optimal signalling strategies by committing to Blackwell experiments (Kamenica and Gentzkow, 2011; Mathevet, Pearce and Stacchetti, 2022; Best and Quigley, 2024; Deb, Pai and Said, 2023) instead of committing directly to HR policies (Milgrom and Roberts, 1988; Waldman, 2003). The result also allows us to relax the commitment assumption using ideas from the information design literature (Lin and Liu, 2024).

1 Set up

A firm employs a unit mass of junior workers.[8] Each worker has a *type*, which we interpret as the worker's ability, and the distribution of workers' types is given by a CDF F that has full support on $\Theta := [0,1]$ and admits a density f. The firm can *let go*, *retain*, or *promote* its workers. We assume that each worker's output is affine and strictly increasing in the worker's type.[9] Letting $j = 1$ and 2 denote the junior and senior positions, respectively, the output of a type-θ worker in position $j \in \{1,2\}$ is given by

$$v_j(\theta) := s_j + k_j\theta,$$

where $s_j \in \mathbb{R}$ and $k_j \in \mathbb{R}_{++}$. The intercept s_j can be interpreted as firm-and-position-specific human capital,[10] and the slope k_j as the marginal product of the worker's ability in position j. Following Costa (1988), we assume that the more senior position has a higher marginal product but not a uniformly higher total product; i.e., $k_2 > k_1$ and $v_2(\theta) \geq v_1(\theta)$ if and only if $\theta \geq \overline{\theta}_1$ for some $\overline{\theta}_1 \in (0,1)$.[11] We say that the firm assigns position $j = 0$ to mean that the firm lets go of the worker; in this case, there is zero output, $v_0(\theta) := 0$.

There is a set I of *outside firms* with heterogeneous production technologies. A worker of type θ working at the outside firm $i \in I$ produces an output of $c^i(\theta)$. For each i, we assume that c^i is affine and strictly increasing in θ. Define $c(\theta) := \sup\{c^i(\theta) : i \in I\}$ to be a type-θ worker's highest output across all outside firms.[12] As a supremum of strictly increasing affine functions, $c(\cdot)$ is strictly increasing and convex. To ease the interpretation and presentation of results, we assume that $c(\cdot)$ is, in fact, strictly convex and differentiable. We refer to the outside firm that maximises the output of a type-θ worker as the *θ-outside firm*. Note that θ-outside firm's production technology, which we denote as $c_\theta(\cdot)$, is given

[8] One could introduce a "pre-stage" game as in Waldman (1984) in which firms that are symmetrically informed about a pool of workers initially compete to hire the workers. Firms subsequently privately learn about the abilities of the workers they hire, and workers gain firm-specific skills as they work. The game we introduce is then a subgame that follows after this pre-stage game for a particular firm.

[9] It is possible to extend our model to allow for non-linear production technologies.

[10] One could introduce a "pre-stage" as in Waldman, 1984 before the start of our game in which firms that are symmetrically informed about a pool of workers initially compete to hire the workers. Firms subsequently privately learn about the abilities of the workers they hire, and workers gain firm-specific skills as they work. The game we study can be thought of as a subgame that follows after this pre-stage game for a particular firm.

[11] More precisely, we assume that $\overline{\theta}_1 \equiv \frac{s_1 - s_2}{k_2 - k_1} \in (0,1)$.

[12] In case an outside firm has multiple positions, we may treat the production technology associated with each position as that of distinct firms.

by the tangent to $c(\cdot)$ at θ, and its worker's marginal product is given by $c'(\theta)$. Thus, the convexity of $c(\cdot)$ reflects that the θ-outside firm's output is more sensitive to workers' abilities when θ is higher—this property will be the key driver of our results.

For the main body of the paper, we further assume $c(\cdot)$ to be such that, for some $\overline{\theta}_0 \in (0, \overline{\theta}_1)$,

$$\max\{c(\theta), v_1(\theta), v_2(\theta)\} = \begin{cases} c(\theta) & \text{if } \theta \in [0, \overline{\theta}_0), \\ v_1(\theta) & \text{if } \theta \in [\overline{\theta}_0, \overline{\theta}_1], \\ v_2(\theta) & \text{if } \theta \in [\overline{\theta}_1, 1]. \end{cases} \tag{1}$$

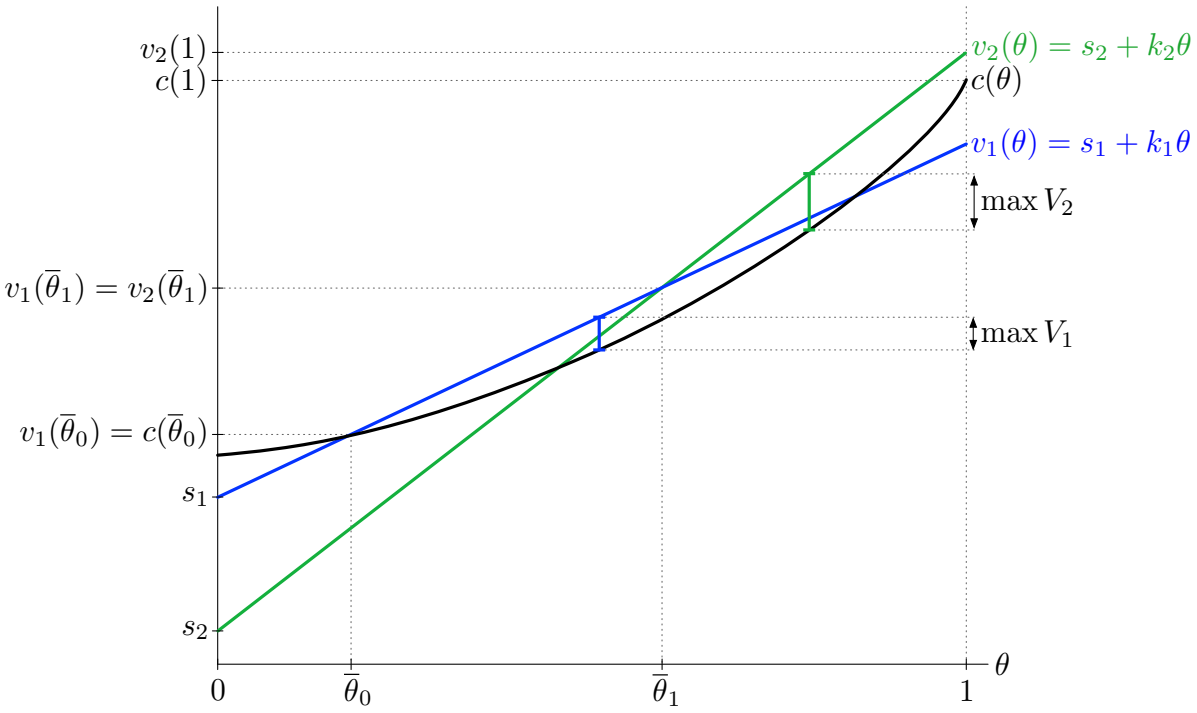
Equation (1) implies that workers whose type is below $\overline{\theta}_0$ are valued by some outside firm more than by the incumbent, and that workers whose type is above $\overline{\theta}_0$ are valued the most by the incumbent firm. This oft-made assumption (e.g., Waldman, 1984; Gibbons and Katz, 1991) ensures that it is possible for the incumbent firm to retain some of its workers and may be justified by factors such as workers' incumbent-firm-specific human capital, switching costs for workers, hiring and training costs for outside firms, or simply by the incumbent firm's superior production technology.[13]

For an allocation of workers to firms to be efficient, each worker must be employed at a firm (and in a position) that maximises the worker's output. Thus, equation (1) means that it is efficient for the firm to let go of workers whose type is below $\overline{\theta}_0$, retain workers whose type lies in $[\overline{\theta}_0, \overline{\theta}_1]$ at the junior level, and to retain and promote workers whose type is above $\overline{\theta}_1$.

A worker's payoff is simply the wage. The firm's payoff from paying a wage w to a type-θ worker in position j is $v_j(\theta) - w$. As we explain in more detail below, we follow Waldman (1984) and think of the firm as paying its workers just enough to ensure that other firms will not poach the workers. Thus, $c(\theta)$ can be interpreted as the wage cost for the firm of a type-θ worker when the labour market can observe θ. We define $V_j \coloneqq v_j - c$ as the firm's profit in such a case and assume that $\max V_2 > \max V_1$ to capture the idea that seniors can potentially be more profitable than junior workers. Figure 1 gives a typical example of the firm's production technology that satisfies our assumptions.

[13] We discuss the case in which outside firms value sufficiently high types of workers more than the incumbent firm in section 5.3.

Figure 1: Value of workers.



1.1 Firm's HR policy design problem

We study the firm's problem of designing an *HR policy*, $\varphi : \Theta \to \Delta(\{0,1,2\})$, that (possibly stochastically) lets go, retains at the current position, or promotes each type of worker. We assume that the market can distinguish between workers based only on their assigned positions. Thus, to retain a worker in position $j \in \{1,2\}$, it suffices for the incumbent employer to just outbid the market and pay a wage equal to the market's expectation of the types of worker in position j.[14] Given a policy φ, the expected type of a worker in position j is given by $\mathbb{E}_\varphi[\theta|j] = \frac{\int_\Theta \theta\varphi(j|\theta)\mathrm{d}F(\theta)}{\int_\Theta \varphi(j|\theta)\mathrm{d}F(\theta)}$. Let $\Psi := (\Delta(\{0,1,2\}))^\Theta$ denote the set of all HR policies. Then, the firm's problem can be written as:

$$\max_{\varphi\in\Psi} \int_\Theta \sum_{j=1}^{2} \left[v_j(\theta) - c\left(\mathbb{E}_\varphi\left[\tilde{\theta}|j\right]\right)\right] \varphi(j|\theta)\,\mathrm{d}F(\theta). \tag{2}$$

[14] We implicitly assume a Bertrand-type competition for the workers so that $c(\cdot)$, strictly speaking, represents the second highest output of workers among firms in the labour market (see, for example, Blume, 2003). We assume that workers break ties in favour of the incumbent employer.

We do not impose incentive compatibility constraints on the firm; the firm is allowed to adopt HR policies that, ex post, result in losses for some types of workers. The lack of such constraints makes the firm's problem distinct from existing promotion-as-signals models and, as we explain in Section 4, allows for the possibility of over-retention and over-promotion by the firm.

2 Solving for the firm's optimal HR policy

The lack of incentive compatibility constraints in the firm's problem means that traditional methods to solve signalling games—establishing that "cutoff" signalling strategies are sufficient and identifying the cutoffs using incentive compatibility constraints—cannot be used to solve the firm's problem. We instead solve the firm's problem indirectly by studying a related information design problem in which, instead of designing an HR policy, the firm designs a public information structure about workers' types. We establish that this information design problem is payoff-equivalent to the firm's HR policy design problem, and that the firm's optimal HR policy can be recovered from the solution to its information design problem.

2.1 Firm's information design problem

Toward describing the firm's information design game, let M denote the set of payoff-irrelevant possible realisations of Blackwell experiments with $|M| \geq 3$, and let $\Pi := (\Delta M)^{\Theta}$ denote the set of possible experiments that the firm can choose. The firm first publicly chooses a Blackwell experiment, $\pi \in \Pi$, about the workers' productivity, the result of which is publicly observed by all firms. We think of such an experiment as a test of whether a worker has reached a particular milestone or passed an exam, and its result as a job title or a qualification. Because the result of the experiment is publicly observed, the firm has no private information about workers' types in this information design game. Consequently, the firm's choice of a Blackwell experiment, π, together with the prior distribution F, completely determines how the firm and the labour market symmetrically learn about the types of workers. Having observed the outcome of the experiment, $m \sim \pi(\cdot|\theta)$, the firm assigns a position j to each worker in a sequentially rational manner. The outside firms, having observed π and (m, j) for each worker, compete over the workers.

We use a belief-based approach to set up the firm's information design problem. Be-

cause production technologies are affine in θ, both the firm's and the market's decisions depend only on the posterior mean after observing (π, m) and (π, m, j), respectively. Moreover, because the firm and the labour market are symmetrically informed about workers' types, the assignment of positions provides no more information than that provided through the result of the experiment. Thus, if the posterior mean of a group of workers is $E \in \Theta$, then the firm must pay $c(E)$ to retain them, regardless of promotion. Consequently, for each realized posterior, the firm assigns it a position that maximizes the expected payoff from those workers. Thus, the firm's payoff from inducing posterior mean E is

$$V(E) \coloneqq \left[\max_{j \in \{1,2\}} v_j(E) - c(E) \right]_+,$$

where $[\cdot]_+ \equiv \max\{\cdot, 0\}$. We refer to $V(\cdot)$ as the firm's value function.

Finally, because of the well-known fact that the firm can induce any distribution of posterior means that is a mean-preserving contraction of F (and the converse is also true), we can write the firm's information design problem as

$$\widehat{V}^* \coloneqq \max_{G \in \mathrm{MPC}(F)} \int_\Theta V(E) \, \mathrm{d}G(E), \tag{3}$$

where $\mathrm{MPC}(F)$ denotes the set of all mean-preserving contractions of F.

The space of mean-preserving contractions of the prior distribution, $\mathrm{MPC}(F)$, is a complex object to analyse. Thus, to proceed, we further simplify the firm's information design problem by showing that it is without loss to only consider distributions of posterior means from a parametric family of CDFs. Towards this goal, let us call a tuple $(\theta_0, \theta_1, E_2) \in \Theta^3$ as *feasible* if $\theta_0 \in [0, \overline{\theta}_0]$, $\theta_1 \in [\theta_0, 1]$, and $\mathbb{E}[\theta | \theta \geq \theta_0] \leq E_2 \leq \mathbb{E}[\theta | \theta \geq \theta_1]$. Given any feasible tuple $(\theta_0, \theta_1, E_2)$, we define $G_{\theta_0, \theta_1, E_2} : \mathbb{R} \to [0,1]$ to be a CDF such that (i) $G \in \mathrm{MPC}(F)$; (ii) the mass on $[0, \theta_0)$ from F is put on $\mathbb{E}[\theta | \theta < \theta_0]$; and (iii) the mass on $[\theta_0, 1]$ from F is put on $\mathbb{E}[\theta | \theta_0 \leq \theta < \theta_1]$ and on E_2. We say that a distribution G is *feasible* if there exists a feasible tuple $(\theta_0, \theta_1, E_2)$ such that $G = G_{\theta_0, \theta_1, E_2}$ and, in that case, we also say that the tuple $(\theta_0, \theta_1, E_2)$ is *associated* with G. The lemma below shows that it suffices to restrict attention to the set of feasible distributions of posterior means.

Lemma 1. *There exists a feasible G^* that solves the information design problem,* (3)*; i.e.,*

$$\widehat{V}^* = \max_{(\theta_0, \theta_1, E_2)} \int_\Theta V(E) \, \mathrm{d}G_{\theta_0, \theta_1, E_2}(\theta) \text{ s.t. } (\theta_0, \theta_1, E_2) \text{ is feasible}. \tag{4}$$

We prove Lemma 1 using results from Arieli et al. (2023). In particular, they show that information design problems, such as the firm's problem, (3), always have a solution that belongs in a family of CDFs that contains the set of feasible CDFs. We then use the structure of V to show that a feasible CDF always solves the firm's problem. Importantly, because (4) is a maximisation problem with respect to three real-valued parameters, Lemma 1 tells us that we can solve the firm's information design problem, (3), and characterise optimal distributions of posterior beliefs using the first-order approach.[15]

2.2 Equivalence of HR policy and information design problems

We now establish an equivalence between the firm's original problem and the information design game.

Lemma 2. *The value of the firm's problem,* (2)*, in the original game equals the value of the firm's problem,* (3)*, in the related game; i.e.,* $V^* = \widehat{V}^*$*. Moreover, if* G^* *solves* (3)*, then there exists an HR policy* $\varphi^* \in \Psi$ *that solves the firm's original problem,* (2)*, and induces* G^* *as the distribution of posterior means.*

That $V^* \leq \widehat{V}^*$ is immediate from the fact that we can always construct a Blackwell experiment, π^*, that induces the same posterior means as an optimal HR policy, φ^*, by setting $\pi^*(m_j|\theta) = \varphi^*(j|\theta)$ for all $\theta \in \Theta$ and all $j \in \{0,1,2\}$ for some $\{m_0, m_1, m_2\} \subseteq M$. To see why $V^* \geq \widehat{V}^*$, fix an optimal experiment π^*. We can then construct another experiment π^{**} by pooling posterior means that induce the firm to assign workers to the same position. Crucially, because each v_j is affine and c is convex, the firm's payoff from π^{**} must be weakly greater than that of π^* so that π^{**} must also be optimal. The proof is completed by observing that we can construct an optimal HR policy φ^{**} by setting $\varphi^{**}(j|\theta) = \pi^{**}(m_j|\theta)$ for all $\theta \in \Theta$ and all $j \in \{0,1,2\}$, where each m_j is associated with a posterior mean that induces the firm to assign position j to workers.

2.3 Monotonicty and determinicity of optimal HR policies

Once we obtain an optimal distribution of posterior means that solve the firm's information design problem, Lemma 2 tells us we can recover an HR policy that solves the original

[15] We note that the objective function need not be concave nor differentiable everywhere so the first-order conditions may not be sufficient for optima. Nevertheless, we are able to solve the problem using the first-order approach by using the special properties of our environment and results from Dworczak and Martini (2019) and Arieli et al. (2023).

problem. However, in general, there is a multiplicity of optimal HR policies because many information structures can induce a particular distribution of posterior means.[16]

Among optimal HR policies, we focus our attention on those with the intuitive property that higher-type workers are more likely to be assigned to a higher position. We refer to such HR policies as being *monotone*.[17] Monotone HR policies have the desirable property that, in an enriched environment in which worker types are endogenous, higher-type workers would not have the incentive to pretend to be lower-type workers (for example, by shirking).[18] As we demonstrate later, it will also turn out to be related to the types of HR policies that can be implemented even when firms do not have "full" commitment power. We now establish that it is without loss to focus on monotone HR policies.

Lemma 3. *Any feasible distribution of posterior means can be induced by a monotone HR policy.*

The proof uses that feasible distributions are special cases of "bi-pooling" distributions of posterior means which Arieli et al. (2023) show can be induced by a monotone information structure. For example, given a feasible tuple $(\theta_0, \theta_1, E_2)$, notice that the following HR policy φ induces $G_{\theta_0,\theta_1,E_2}$: φ lets go of the workers whose types lie in $[0,\theta_0]$ with probability one, promotes workers whose types lie in $[\theta_1,1]$ with probability one, and randomise between promoting and not promoting (but retaining) workers whose types lie in $[\theta_0,\theta_1]$. Thus, we interpret θ_0 as a *retention threshold* and θ_1 as a *always-promote threshold*, where we make the distinction to account for the possible randomisation.

As the example above suggests, not every feasible distribution of posterior means can be induced by a deterministic HR policy that assigns each type to a position with probability one. In particular, no deterministic HR policy can induce feasible distributions of

[16] See Example 3 in Arieli et al. (2023).

[17] To define this property formally, note that an HR policy φ, when viewed as an information structure (i.e., with $M = \{0,1,2\}$ and $\varphi = \pi$), induces three possible posterior means, $\{E_j\}_{j=0}^2$, where E_j is the posterior mean after observing $j \in M$. Let $\widehat{\varphi}(\theta) \in \Delta(\{E_j\}_{j=0}^2)$ denote the distribution of posterior means such that

$$\widehat{\varphi}(E|\theta) = \begin{cases} \varphi(j|\theta) & \text{if } \exists j \in \{0,1,2\},\ E = E_j, \\ 0 & \text{otherwise.} \end{cases}$$

With this notation, an HR policy φ is *monotone* if, for any $\theta' \geq \theta$, $\widehat{\varphi}(\theta')$ first-order stochastically dominates $\widehat{\varphi}(\theta)$. Observe that if φ is deterministic, then the support of $\widehat{\varphi}(\cdot)$ is a singleton set, and thus, monotonicity reduces to requiring posterior means to be higher for higher types of workers.

[18] Goldstein and Leitner (2018) and Mensch (2021) provide discussions on why and when the monotonicity property might be desirable.

posterior means associated with E_2 that lie in the interval $(\mathbb{E}[\theta|\theta \geq \theta_0], \mathbb{E}[\theta|\theta \geq \theta_1])$.[19] The proposition below fully characterises the cases in which a deterministic HR policy is optimal.

Proposition 1. *A deterministic HR policy is optimal if and only if there does not exist a tuple* $(\theta_0^b, E_1^b, E_2^b)$ *with* $\theta_0^b \leq E_1^b \leq E_2^b$ *such that*

$$c\left(E_j^b\right) - v_j\left(\theta_0^b\right) = c'\left(E_j^b\right)\left(E_j^b - \theta_0^b\right) \ \forall j = \{1,2\}, \tag{5}$$

$$k_2 - k_1 = c'\left(E_2^b\right) - c'\left(E_1^b\right), \tag{6}$$

$$E_1^b \leq \mathbb{E}\left[\theta|\theta \geq \theta_0^b\right] < E_2^b < \mathbb{E}\left[\theta|\theta \geq \theta_1^b\right], \tag{7}$$

where θ_1^b *satisfies* $\mathbb{E}[\theta|\theta_0^b \leq \theta < \theta_1^b] = E_1^b$.

To understand the proposition, observe that existence of a tuple $(\theta_0^b, E_1^b, E_2^b)$ satisfying condition (ii) of Proposition 1 implies existence of a feasible CDF associated with a tuple $(\theta_0^b, \theta_1^b, E_2^b)$ where $\mathbb{E}[\theta|\theta \geq \theta_0^b] < E_2^b < \mathbb{E}[\theta|\theta \geq \theta_1^b]$ so that a deterministic HR policy cannot induce it. Roughly, condition (7) requires a sufficient mass of workers who are "good enough" so that the firm can reduce the wage cost of senior workers by sometimes promoting workers who are not "good enough." Conditions (5) and (6), which are restrictions on v_1, v_2 and c but not on F, ensure that such a feasible CDF is optimal in the firm's information design problem. Specifically, condition (5) ensures that setting the retention threshold at θ_0^b is optimal for the firm that chooses an always-promote threshold of θ_1^b.[20] Condition (6), which requires the firm and the outside firms to value the additional marginal product from senior over junior workers equally, ensures that setting the always-promote threshold at θ_1^b is optimal when inducing posterior means for junior and senior workers of E_1^b and E_2^b, respectively.

The discussion above implies that we can guarantee the existence of an optimal deterministic HR policy when: either no tuple exists that satisfies condition (i) of Proposition 1; or, even if such a tuple exists, the prior distribution F is such that conditions (ii) does not hold. In the appendix, we show that for a tuple $(\theta_0^b, E_1^b, E_2^b)$ satisfying conditions (i) of

[19] The case in which $E_2 = \mathbb{E}[\theta|\theta \geq \theta_0]$ is the case in which all workers above the retention threshold are promoted (i.e., $\theta_0 = \theta_1$). In contrast, $E_2 = \mathbb{E}[\theta|\theta \geq \theta_1]$ is the case in which workers are promoted if and only if their type is above θ_1 so that θ_1 can be interpreted simply as the promotion threshold.

[20] For each $j \in \{1,2\}$, the condition (5) ensures that setting the retention threshold at θ_0^b is optimal for a firm that chooses between letting go of workers and retaining workers in position j while inducing a posterior mean of E_j^b.

Proposition 1 to exist, $c(\cdot)$ must be sufficiently convex but not too convex. Thus, to benefit from randomisation, the production technologies of outside firms cannot vary too widely.

3 Inefficiencies of optimal HR policies

Our main results concern the nature of inefficiencies associated with the firm's optimal retention and promotion decisions.[21] To be precise, we say that an HR policy φ exhibits *over-retention* (resp. *under-retention*) if φ retains workers whose types are below (resp. above) the efficient retention cutoff, $\overline{\theta}_0$, with positive probability.[22] Under this definition, an HR policy exhibits efficient retention if and only if it does not exhibit over-retention or under-retention. We emphasise that an HR policy can exhibit both over- and under-retention if it over-retains some types of workers and under-retains other types of workers. Analogously, we say that φ exhibits *over-promotion* (resp. *under-promotion*) if φ promotes workers whose types are below (resp. above) the efficient promotion cutoff, $\overline{\theta}_1$, with positive probability.

3.1 Over-retention is optimal

We first establish that over-retention is optimal for the firm except in a knife-edge case where the optimal policy retains workers efficiently.

Theorem 1. *An optimal HR policy exists with efficient retention if and only if conditions*

$$V_2'(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0]) \geq V_1'(\overline{\theta}_0) \tag{8}$$

and

$$V_2(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0]) = V_2'(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0])(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0] - \overline{\theta}_0) \tag{9}$$

hold. Moreover, under-retention is never optimal. Therefore, if conditions (8) *or* (9) *do not hold, then any optimal HR policy exhibits over-retention.*

[21] In the appendix, we provide a full characterisation of the firm's optimal monotone HR policies.

[22] We can equivalently interpret there to be a continuum of workers of each type such that the distribution of types corresponds to F. Under such an interpretation, φ can be thought of as a deterministic mapping from $(i,\theta) \in [0,1] \times \Theta$ to a position in $j \in \{0,1,2\}$.

We emphasise that conditions (8) and (9) hold only in knife-edge cases.[23] Thus, we interpret Theorem 1 as saying that we should generally expect firms to over-retain workers.

The underlying intuition for Theorem 1 is most clearly demonstrated by considering a cutoff HR policy in which the firm retains and promotes a worker if and only if the worker's type is above a retention cutoff θ_0 and a promotion cutoff $\theta_1 > \theta_0$, respectively. Such a cutoff policy can be implemented with an information structure that separately pools intervals $[0, \theta_0)$, $[\theta_0, \theta_1)$, and $[\theta_1, 1]$. The firm's profits from the cutoff policy is given by

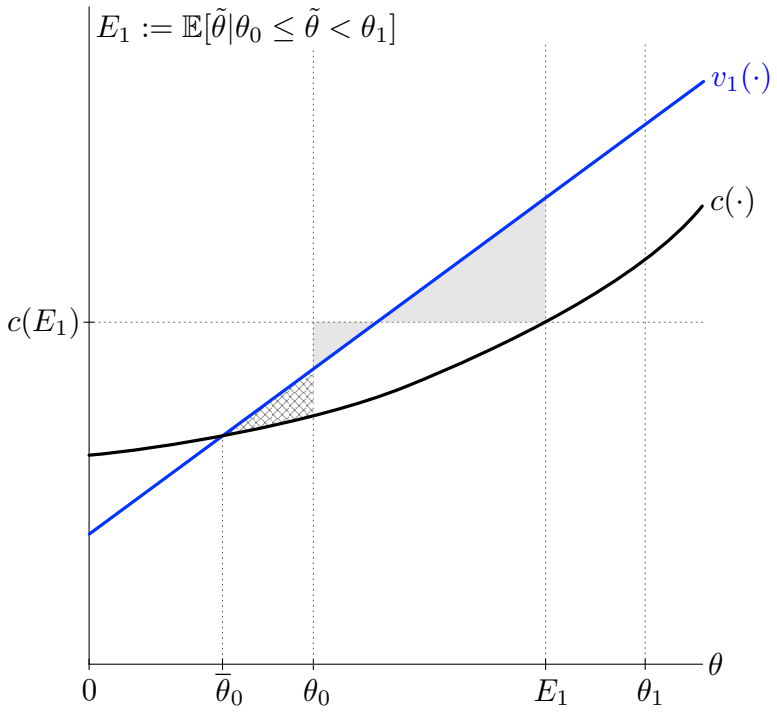
$$\begin{aligned}\widehat{V}(\theta_0, \theta_1) = &\int_{\theta_0}^{\theta_1} \left(v_1(\theta) - c\left(\mathbb{E}\left[\tilde{\theta} | \theta_0 \leq \tilde{\theta} < \theta_1\right]\right)\right) f(\theta)\, \mathrm{d}\theta \\ &+ \int_{\theta_1}^{1} \left(v_2(\theta) - c\left(\mathbb{E}\left[\tilde{\theta} | \tilde{\theta} \geq \theta_1\right]\right)\right) f(\theta)\, \mathrm{d}\theta, \end{aligned} \tag{10}$$

where the first and second lines are profits associated with workers in the junior and senior positions, respectively.

Let us first understand why under-retention is never optimal. Figure 2 below shows the case in which the firm adopts a retention cutoff that is strictly above the efficient retention cutoff, $\theta_0 > \overline{\theta}_0$. The two grey-shaded triangles in the figure represent the profits from each type of worker in the junior position under such a policy.

[23] In words, the conditions require the tangent to V at $\mathbb{E}[\theta | \theta \geq \overline{\theta}_0]$ to have a slope greater than $V_1'(\overline{\theta}_0)$ and to cut the x-axis at exactly $\overline{\theta}_0$.

Figure 2: Under-retention is not optimal.



Now consider modifying the information structure associated with the cutoff HR policy so that it fully reveals the types of workers in the interval $(\overline{\theta}_0, \theta_0)$. With this modification, the firm obtains strictly profits, $v_1(\theta) - c(\theta) > 0$, for all types of workers in the interval $(\overline{\theta}_0, \theta_0)$—the hatched area in the figure above—while profits associated with other types of workers remain unchanged. Since the firm can always make such a strictly profitable modification to any information structure associated with an HR policy that exhibits under-retention, it follows that under-retention is never optimal for the firm.[24] In other words, under-retention is suboptimal because it implies the firm "gives up" workers even though it values the workers more than any outside firm.

To see why over-retention is generally optimal, consider the firm's incentive to "just" over-retain; i.e., to choose a retention cut-off, θ_0, just below the efficient cutoff $\overline{\theta}_0$. We can do so by analysing the derivative of $\widehat{V}_1(\theta_0, \theta_1)$ with respect to θ_0 evaluated at the efficient retention cutoff $\overline{\theta}_0$ while assuming that $E_1 \equiv \mathbb{E}[\theta|\overline{\theta}_0 \leq \theta < \theta_1] \in (\overline{\theta}_0, \overline{\theta}_1)$ (which can be

[24] We note that the modified information structure is itself not optimal for the firm.

shown to hold in the relevant cases). Leibniz rule gives that

$$\left.\frac{\partial \widehat{V}(\cdot)}{\partial \theta_0}\right|_{\theta=\overline{\theta}_0} = -\left[f\left(\overline{\theta}_0\right)\left(v_1\left(\overline{\theta}_0\right)-c\left(E_1\right)\right)+\int_{\overline{\theta}_0}^{\theta_1} c'\left(E_1\right)\frac{\partial E_1}{\partial \theta_0} f(\theta)\mathrm{d}\theta\right], \tag{11}$$

The first term (inside the square brackets) in (11) is the direct effect on the firm's profit from letting go of a type-$\overline{\theta}_0$ worker: the firm pays $c(E_1)$ to retain a type-$\overline{\theta}_0$ worker that produces $v_1(\overline{\theta}_0)$ in output. The second term reflects the signalling effect of a change in the retention threshold on the firm's wage cost: $\frac{\partial E_1}{\partial \theta_0}$ captures the changes in the market's expectation about types of junior workers, multiplying it by $c'(E_1)$ gives the highest expected output any outside firm can produce with a junior worker from the incumbent firm,[25] and the integral sums the change in the wage rate associated with all junior workers.

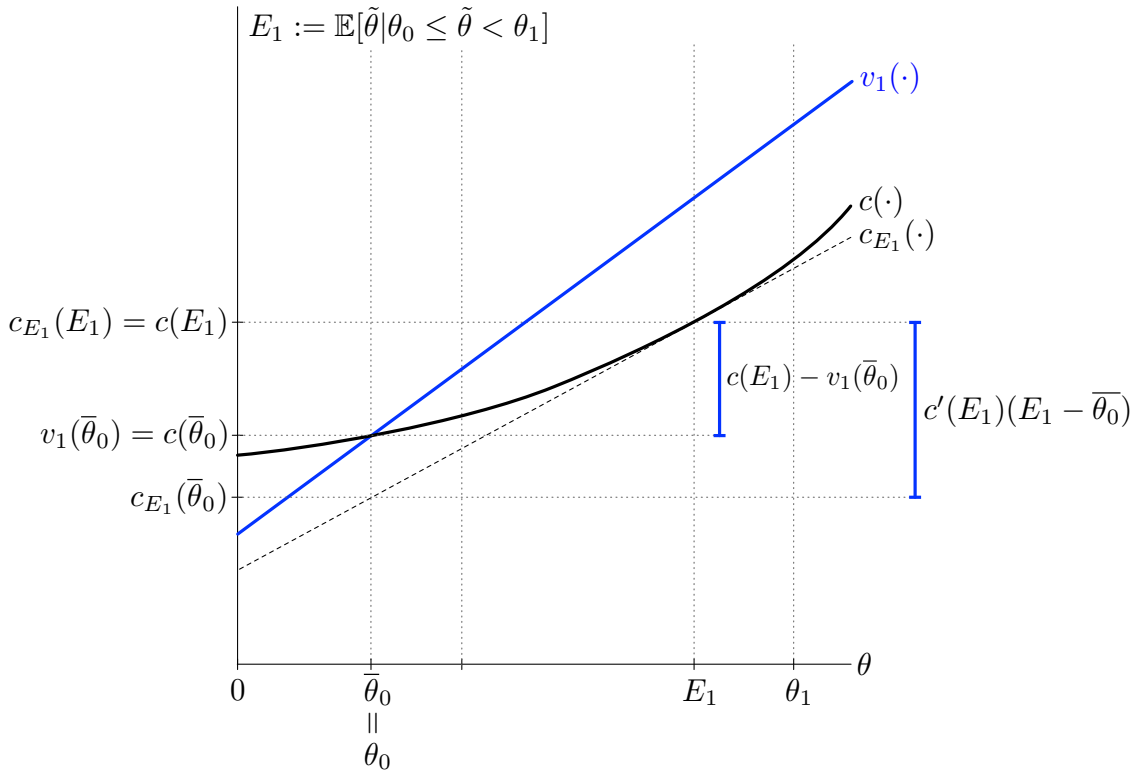
The first term in (11) is strictly negative because the labour market's willingness to pay for a worker is increasing in the worker's type (i.e., c is strictly increasing) and a type-$\overline{\theta}_0$ worker, by definition, is equally valued by the firm and the labour market. The second term, on the other hand, is strictly positive because the posterior mean of junior workers is increasing in retention cutoff, and the labour market has a positive willingness to pay for junior workers. Thus, a priori the sign of the derivative seems ambiguous. However, it turns out that the convexity of c (which arises from the fact that θ-outside firm's technology is more sensitive to workers' abilities when θ is higher) implies that the signalling effect dominates the direct effect on firm profits. This can be seen by rewriting the derivative (11) as follows:

$$\begin{aligned}\left.\frac{\partial \widehat{V}(\cdot)}{\partial \theta_0}\right|_{\theta=\overline{\theta}_0} &= f\left(\overline{\theta}_0\right)\left[\left(c\left(E_1\right)-v_1\left(\overline{\theta}_0\right)\right)-c'\left(E_1\right)\left(E_1-\overline{\theta}_0\right)\right] \\ &= f\left(\overline{\theta}_0\right)\left[\left(c\left(E_1\right)-c\left(\overline{\theta}_0\right)\right)-\left(c_{E_1}\left(E_1\right)-c_{E_1}\left(\overline{\theta}_0\right)\right)\right], \end{aligned} \tag{12}$$

where the first line is obtained by substituting the expression for $\frac{\partial E_1}{\partial \theta_0}$, and the second line uses the fact that $c(\overline{\theta}_0) = v_1(\overline{\theta}_0)$ by definition of $\overline{\theta}_0$ and $c'(E_1)$ is the slope of the production technology of the E_1-outside firm, c_{E_1}. As evident from Figure 3, the convexity of c implies that the second difference in (12) is strictly greater than the first. Consequently, the derivative, (11), is strictly negative meaning that the firm has a strict incentive to over-retain.

[25] To see this, observe that $c'(E_1) \equiv c_{E_1}(E_1)$, where $c_{E_1}(\cdot)$ is the production technology associated with E_1-outside firm.

Figure 3: Over-retention.



In words, the signalling effect dominates the direct effect because the wage cost is determined by the E_1-outside firm that values all junior workers at the "higher" average marginal product rather than at their actual marginal product. Hence, the firm strictly benefits from over-retaining workers over efficiently retaining workers.

In proving Theorem 1, we also show that there is over-retention even when the optimal G is associated with $(\theta_0, \theta_1, E_2)$ for which $E_2 < \mathbb{E}[\theta|\theta \geq \theta_1]$ (which includes cases in which optimal HR policy is not deterministic). In this case, the effect of changing the retention threshold is more complex, because changing the retention threshold also affects the mass of workers in position $j = 2$.[26] Nevertheless, the same forces described above lead the firm to over-retain workers.

[26] To see why, in general, the firm's expected payoff from a feasible G associated with $(\theta_0, \theta_1, E_2)$ is

$$(1 - F(\theta_0))\left[V(\mathbb{E}[\theta|\theta_0 \leq \theta < \theta_1])\,\alpha(\theta_0, \theta_1, E_2) + V(\mathbb{E}[\theta|\theta \geq \theta_1])(1 - \alpha(\theta_0, \theta_1, E_2))\right],$$

where $\alpha(\theta_0, \theta_1, E_2)$ represents the proportion of retained workers who are assigned to position $j = 1$. Hence, changing the retention threshold θ_0 now also affects the mass of workers who are retained in position $j = 2$.

3.2 Over-promotion is sometimes optimal

We now ask: Does the firm's optimal promotion decision also always lead to over-promotion? Even if we focus on optimal monotone policies, it turns out that the nature of the inefficiency associated with promotion decisions is more nuanced.

Theorem 2. *An optimal deterministic monotone HR policy exhibits either over-promotion or under-promotion but not both. If no deterministic monotone HR policy is optimal, then an optimal monotone HR policy could exhibit only over-promotion, only under-promotion, or both.*

Once again, the intuition for why the results differ between retention and promotion decisions is most clear when considering a deterministic monotone HR policy. Thus, consider again the firm's expected payoff from such a policy given by equation (10). Observe that a change in the always-promote threshold θ_1 affects not only senior workers (i.e., $j = 2$) but also junior workers (i.e., $j = 1$). In contrast, changing the retention threshold only affects the junior workers because the firm's average profit-per-worker from letting go of workers (i.e., $j = 0$) is unaffected by its retention decisions—it always equals zero. Hence, compared to the case of over-retention, there is an additional force to consider in the case of over-promotion.

To be more precise, let us consider the derivative of $\widehat{V}$ with respect to θ_1 evaluated at the efficient promotion threshold $\overline{\theta}_1$ assuming that $E_1 \equiv \mathbb{E}[\theta|\theta_0 \leq \theta < \overline{\theta}_1] \in (\overline{\theta}_0, \overline{\theta}_1)$ and $E_2 \equiv \mathbb{E}[\theta|\theta \geq \overline{\theta}_1] \in (\overline{\theta}_1, 1)$ (which can be shown to hold in the relevant cases):

$$\begin{aligned}\left.\frac{\partial \widehat{V}(\cdot)}{\partial \theta_1}\right|_{\theta=\overline{\theta}_1} &= f\left(\overline{\theta}_1\right)\left[\left(v_1\left(\overline{\theta}_1\right) - c\left(E_1\right)\right) - \left(c_{E_1}\left(\overline{\theta}_1\right) - c_{E_1}\left(E_1\right)\right)\right], \\ &\quad + f\left(\overline{\theta}_1\right)\left[\left(c\left(E_2\right) - v_2\left(\overline{\theta}_1\right)\right) - \left(c_{E_2}\left(E_2\right) - c_{E_2}\left(\overline{\theta}_1\right)\right)\right], \qquad (13)\end{aligned}$$

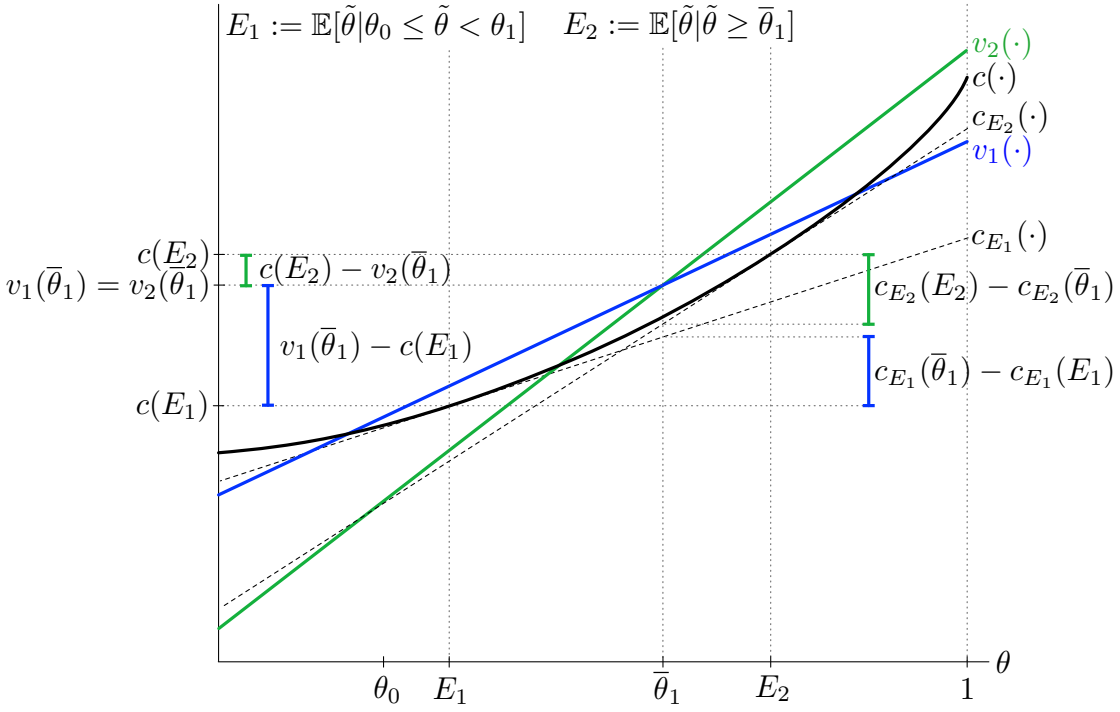
where the first and second lines are changes to the firm's profits from junior and senior workers, respectively.

Now, suppose the firm slightly over-promotes by promoting a worker whose type is just below $\overline{\theta}_1$. Although such over-promotion increases the firm's total output from senior workers, the increase is exactly cancelled out by the reduction in total output from junior workers because the firm's net gain, $v_2(\overline{\theta}_1) - v_1(\overline{\theta}_1)$, equals zero by the definition of the efficient promotion cutoff $\overline{\theta}_1$. However, the promoted marginal worker, who was previously paid $c(E_1)$, must now be paid $c(E_2)$. Hence, over-promotion is profitable for the firm

only if the signalling effects are greater than the increased wage cost $c(E_2) - c(E_1) > 0$.

Over-promotion reduces wage rates for workers in both junior and senior positions,[27] and as in the case of retention, the magnitude of the reduction depends on the production technologies of E_1- and E_2-outside firms. The convexity of c is sufficient to ensure that the second line in (13) is strictly negative.[28] However, convexity is not sufficient to determine the sign of the first term, which depends on the relative marginal products of workers at E_1- and E_2-outside firms.

Figure 4: Under-promotion is optimal.



In particular, if the marginal product of the E_1-outside firm is low compared to that of the E_2-outside firm, the increased wage cost associated with the marginal worker can dominate the reduction in wage rates from signalling effects. This case is shown in Figure 4 above in which the increase in wage cost from over-promotion (the sum of the vertical bars to the left) is greater than the benefit from the signalling effect (the sum of the vertical vars to the right). In this case, the firm prefers to under-promote instead of over-promote workers. Alternatively, it is also possible that the marginal product of workers is relatively

[27] Recall $E_1 \equiv \mathbb{E}[\theta|\theta_0 \le \theta < \overline{\theta}_1]$ and $E_2 \equiv \mathbb{E}[\theta|\theta \ge \overline{\theta}_1]]$ so that they are both decreasing in $\overline{\theta}_1$.

[28] We note that strict convexity of c is not necessary to sign the second line.

high at the E_1-outside firm so that the signalling effects dominate the increased wage cost from over-promotion so that the firm would find it profitable to over-promote workers.

In the case where no deterministic HR policy is optimal, an optimal monotone policy exists with a retention threshold of $\theta_0^* \leq \overline{\theta}_0$ and an always-promotion threshold of $\theta_1^* > \theta_0^*$. If $\theta_1^* \leq \overline{\theta}_1$, such an HR policy over-promotes workers because workers of types in the interval $[\theta_0^*, \theta_1^*)$ are promoted with positive probability and workers in the interval $[\theta_1^*, \overline{\theta}_1]$ are promoted with probability one. If, instead, $\theta_1^* > \overline{\theta}_1$, then the policy both over- and under-promotes workers because it promotes workers of types in the interval $[\theta_0^*, \theta_1^*)$ with positive probability but doesn't promote workers of types in the interval $[\overline{\theta}_1, \theta_1^*)$ with probability one.

3.3 When does the firm over-promote?

We focus on the case in which an optimal monotone HR policy is deterministic so that the optimal monotone HR policy either exhibits over-promotion or under-promotion but not both. To state when the firm would over- or under-promote, let $\overline{E}_1^e$ denote the expected type of junior workers that makes it optimal for the firm to choose the efficient promotion threshold of $\overline{\theta}_1$, and let $\overline{\theta}_0^e$ denote the optimal retention threshold in case junior worker's expected type is $\overline{E}_1^e$.[29] We note that $\overline{E}_1^e$ and $\overline{\theta}_0^e$ need not be consistent with F so that, in general, we would expect $\overline{E}_1^e \neq \mathbb{E}[\theta | \overline{\theta}_0^e \leq \theta < \overline{\theta}_1]$.

Proposition 2. *Suppose a deterministic optimal HR policy φ^* exists.*

(i) Suppose there exists a tuple $(\theta_0^b, E_1^b, E_2^b)$ with $\theta_0^b \leq E_1^b \leq E_2^b$ satisfying conditions (5) *and* (6)*. Let θ_1^b be such that $\mathbb{E}[\theta | \theta_0^b \leq \theta < \theta_1^b] = E_1^b$. Then, φ^* exhibits over-promotion if $E_2^b < \mathbb{E}[\theta | \theta \geq \theta_1^b]$ and either (a) $\mathbb{E}[\theta | \theta \geq \overline{\theta}_1] > E_2^b$ or (b) $\mathbb{E}[\theta | \theta \geq \overline{\theta}_1] \leq E_2^b$ and $\mathbb{E}[\theta | \overline{\theta}_0^e \leq \theta < \overline{\theta}_1] > \overline{E}_1^e$; otherwise, φ^* exhibits under-promotion.*

(ii) Suppose no such tuple exists. Then, φ^* exhibits over-promotion if $\mathbb{E}[\theta | \overline{\theta}_0^e \leq \theta < \overline{\theta}_1] > \overline{E}_1^e$; otherwise, φ^* exhibits under-promotion.

Recall that conditions (5) and (6) are restrictions on v_1, v_2 and c but not on F, and that a tuple satisfying these conditions exists if c is convex but not too convex; in other words, when the production technologies of outside firms are not too varied. Thus, case

[29] Concretely, $\overline{E}_1^e$ sets the derivative $\partial \widehat{V} / \partial \theta_1$ evaluated at $\theta = \overline{\theta}_1$ to zero when expected typea of junior and senior workers are $\overline{E}_1^e$ and $\mathbb{E}[\theta | \theta \geq \overline{\theta}_1]$ respectively, and $\overline{\theta}_0^e$ sets the derivative $\partial \widehat{V} / \partial \theta_0$ evaluated at $\theta = \overline{\theta}_0^e$ to zero when expected type of junior workers is $\overline{E}_1^e$.

(ii) in Proposition 2 above says that the firm would over-promote if production technologies of outside firms are sufficiently varied and the expected type of junior workers if the firm were to follow a threshold HR policy $(\overline{\theta}_0^e, \overline{\theta}_1)$ is sufficiently high. The latter condition ensures that the effect of lowering the promotion threshold on the profits from workers in the junior position is sufficiently small. Case (i) in Proposition 2 concerns the environment in which the production technologies of outside firms are not too varied. Here, over-promotion occurs if there is a sufficiently large mass of high-type workers (so that $\mathbb{E}[\theta|\theta \geq \theta_1^b], \mathbb{E}[\theta|\theta \geq \overline{\theta}_1] > E_2^b$) which, in particular, implies promoting efficiently results in wage costs that are "too high" for the firm, Alternatively, over-promotion can occur if, as in case (i), the expected type of junior workers if the firm were to follow a threshold HR policy $(\overline{\theta}_0^e, \overline{\theta}_1)$ is sufficiently high.

To better understand when a firm over-promotes, let us study the firm's incentive to over-promote as we change the environment. To that end, we say that tuples of primitives (v_2, c, F) and $(\tilde{v}_2, \tilde{c}, \tilde{F})$ (satisfying our assumptions) are *comparable* if they both imply a common efficient retention and promotion thresholds, and if they both satisfy conditions for either case (i)(b) or (ii) of Proposition 2. Given such comparable tuples of primitives, we say that over-promotion is *more likely* under (v_2, c, F) than under $(\tilde{v}_2, \tilde{c}, \tilde{F})$ if the change can result in a switch from the optimal monotone HR policy exhibiting under-promotion to over-promotion but not the other way around.[30]

We first contemplate a change in the production technologies of outside firms. Specifically, we consider making comparable changes that alter $c'(\overline{E}_2)$, where $\overline{E}_2 \coloneqq \mathbb{E}[\theta|\theta \geq \overline{\theta}_1]$. To isolate the signalling effect, we ensure $c(\overline{E}_2)$ and $c(1)$ remain unchanged so that the change does not directly alter the wage cost for the firm from promoting efficiently, and we also leave c unchanged on $[0, \overline{\theta}_1]$ to avoid directly affecting the firm's choice of retention threshold.

Proposition 3. *Suppose no deterministic monotone HR policy is optimal. Consider comparable* (v_2, c, F) *and* $(v_2, \tilde{c}, F)$ *such that* $c(\theta) = \tilde{c}(\theta)$ *for* $\theta \in [0, \overline{\theta}_1] \cup \{\overline{E}_2, 1\}$. *Then, over-promotion is more likely under* $(v_2, \tilde{c}, F)$ *than under* (v_2, c, F) *if* $\tilde{c}'(\overline{E}_2) > c'(\overline{E}_2)$.

In words, the above proposition says over-promotion is more likely with a greater mar-

[30] More concretely, we say that over-promotion is more likely under (v_2, c, F) than under $(\tilde{v}_2, \tilde{c}, \tilde{F})$ if the derivative $\frac{\partial \widehat{V}}{\partial \theta_1}$ evaluated at $(\overline{\theta}_0^e, \overline{\theta}_1)$ is smaller under (v, c, F) than under $(\tilde{v}_2, \tilde{c}, \tilde{F})$, where $\overline{\theta}_0^e$ is such that the first-order condition with respect to θ_0 holds. That is, $\frac{\partial \widehat{V}}{\partial \theta_0}$ evaluated at $(\overline{\theta}_0^e, \overline{\theta}_1)$ is zero if $\overline{\theta}_0^e > 0$ or strictly negative if $\overline{\theta}_0^e = 0$. Observe that there cannot be a switch from under- to over-promotion between two tuples of primitives that satisfy condition (i) of Proposition 2.

ginal product of $\overline{E}_2$-outside firm (i.e., the firm that values efficiently promoted workers the highest). The result follows from the fact that the greater marginal product of such a firm leads to a greater magnification of the signalling effect associated with senior position.

To demonstrate that the incentive to over- or unde-promote is largely driven by the technologies of outside firms, we also consider a change in the marginal product of the incumbent firm's production technology, k_2 (with a corresponding change in s_2 to ensure that the efficient promotion threshold remains unchanged). The following proposition shows that such a comparable change does not affect the firm's incentive to over- or under-promote.

Proposition 4. *Suppose no deterministic monotone HR policy is optimal. Consider comparable* (v_2, c, F) *and* $(\tilde{v}_2, c, F)$*. Then, over-promotion is no more likely under* $(\tilde{v}_2, c, F)$ *than under* (v_2, c, F)*.*

We interpret Propositions 3 and 4 as reinforcing the insight that whether there is over-promotion or under-promotion is driven by the technologies of outside firms, rather than the technology of the incumbent employer.

We can also use our model to study how the nature of inefficiency associated with optimal HR policies changes when the workers have different characteristics that are publicly observable; e.g., workers may have different education levels or professional qualifications. To that end, suppose there are two sets of workers and their prior distribution of types is ordered according to their likelihood ratios.[31] We show that over-promotion is more likely for the set of workers whose distribution dominates another in the likelihood-ratio order.

Proposition 5. *Suppose no deterministic monotone HR policy is optimal. Consider comparable* (v_2, c, F) *and* $(v_2, c, \tilde{F})$*. Then, over-promotion is more likely under* $(v_2, c, \tilde{F})$ *than under* (v_2, c, F) *if* $\tilde{F}$ *dominates* F *in the likelihood ratio order.*

For example, if workers with a higher education level are "better" than those with a lower education level in the sense of likelihood ratio order, over-promotion is more likely among workers who are more educated. This observation is consistent with DeVaro and Waldman (2012) who empirically find that, fixing ability, workers with higher education levels are more likely to be promoted.

[31] Recall that, given two absolutely continuous CDFs with support $[0,1]$, F and H, the CDF H dominates F in the likelihood ratio order if and only if $\frac{h(\cdot)}{f(\cdot)}$ is increasing.

4 Relaxing the firm's commitment power

The firm can directly commit to HR policies in a number of ways. For example, the firm can delegate its retention and promotion decisions to an HR department (Milgrom and Roberts, 1988) or establish reputation (Waldman, 2003). The equivalence between the original signalling game and the information design game implies that the firm can also indirectly implement optimal HR policies by committing to public Blackwell experiments about its workers.

For example, the firm could establish a clear career path for workers in which progress depends on the workers reaching certain milestones (or passing some tests) as a way to implement optimal HR policies. Indeed, in many industries, firms disclose to potential workers how they can progress through the ranks.[32] Alternatively, the firm could also implement optimal experiments by coming to contracts à la Deb, Pai and Said (2023).[33] In this light, the resistance from HR departments that line managers often face in trying to get their managees promoted "faster" or firing managees can be viewed as an optimal response for the firm to manipulate the signalling effects of promotion. Nevertheless one may wonder: To what extent do our results depend on the firm's ability to commit?

To further demonstrate how else the firm can implement optimal HR policies, we consider the case in which the firm lacks the power to commit to HR policies, but the labour market can observe the distribution of positions held by its workers after it has assigned them. That is, in addition to observing the firm's choice of HR policy φ and the position of the workers $j \sim \varphi(\cdot|\theta)$, the labour market also observes the distribution of positions after the firm has made its retention and promotion decisions. Firms sometimes disclose such information voluntarily,[34] or disclose them as mandated by accounting rules or laws,[35] alternatively, they are available publicly through websites such as LinkedIn or through ser-

[32] Career fairs at colleges and universities are ripe with employers explaining how graduates can expect to progress through the rank in the company.

[33] The Principal, the Agent, and the Receiver, in their setting, correspond to the firm, a line manager, and the labour market in our setting, where the line manager is an agent of the firm who is able to observe the workers' types. We would add to the set-up of Deb, Pai and Said (2023) that the firm is able to assign positions once it has gathered information about workers from the line manager.

[34] For example, certain types of companies and institutions, such as economic consultancies and academic departments, provide a public list of their staff in various positions.

[35] For example, in the UK, the Companies Act 2006 (section 411) specifies that firms above certain size "must also disclose the average number of persons within each category of persons so employed." Some diversity, equity and inclusion reporting requirements stipulate companies to disclose demographic information about their employees (e.g., The European Union's Directive 2014/95/EU).

vices that provide intelligence about HR structures of companies.[36]

Importantly, public knowledge of the distribution of assigned positions effectively allows the market to detect (and respond to) deviations by the firm from its stated HR policy, φ, if deviations result in a distribution of positions that is inconsistent with φ. Following arguments from Lin and Liu (2024), the possibility for the market to respond to deviations allows the firm to effectively commit to HR policies that result in the same marginal distribution of messages as that of the announced one.

To state the firm's problem when the firm has this partial commitment power, let $D(\varphi) \subseteq \Psi$ denote the set of HR policies that have the same marginal distribution as $\varphi \in \Psi$; i.e.,

$$D(\varphi) \coloneqq \left\{ \varphi' \in \Psi : \int_\Theta \varphi'(\cdot|\theta)\, \mathrm{d}F(\theta) = \int_\Theta \varphi(\cdot|\theta)\, \mathrm{d}F(\theta) \right\}.$$

The firm's problem is to choose an HR policy that maximises its ex ante expected payoff:

$$\max_{\varphi \in \Psi} \int_\Theta \sum_{j=1}^{2} \left[v_j(\theta) - c\left(\mathbb{E}_\varphi \left[\tilde{\theta} | j \right] \right) \right] \varphi(j|\theta)\, \mathrm{d}F(\theta) \tag{14}$$

$$\text{s.t. } \varphi \in \arg\max_{\varphi' \in D(\varphi)} \int_\Theta \sum_{j=1}^{2} \left[v_j(\theta) - c\left(\mathbb{E}_\varphi \left[\tilde{\theta} | j \right] \right) \right] \varphi'(j|\theta)\, \mathrm{d}F(\theta). \tag{15}$$

The constraint (15), which we refer to as the firm's *credibility constraint*, ensures that the firm, after announcing that it will follow an HR policy φ, would not deviate to some other HR policy that has the same marginal distribution of positions when the labour market uses φ to update beliefs about workers. We say that an HR policy φ is *credible* if φ satisfies the credibility constraint.

In solving (14), we first show that we can restrict attention to deterministic and monotone HR policies.

Lemma 4. *An HR policy is credible if and only if it is deterministic and monotone.*

An immediate corollary is that the ability to partially commit is sufficient for the firm to attain the full-commitment payoff whenever there exists a deterministic optimal HR policy exist.

Corollary 1. *Optimal HR policy can be implemented under partial commitment if and only if a deterministic optimal HR policy exists.*

[36] For example, companies such as Crayon collect "Competitive Intelligence" to "provide the most complete picture of what your competitors are up to."

Note that an HR policy φ that is both deterministic and monotone is a simple threshold policy. Thus, Lemma 4 simplifies the problem (2) to one of finding optimal retention and promotion thresholds. Using this simplification, we prove that the nature of inefficiency remains unchanged.

Proposition 6. *Any HR policy φ^* that solves* (14) *is a threshold policy with retention and promotion thresholds, θ_0^* and $\theta_1^* \geq \theta_0^*$ respectively. In addition, φ^* exhibits over-retention, and it may exhibit over- or under-promotion but not both.*

We note that some direct or indirect ability to commit to HR policies is necessary to ensure that firm can implement HR policy that exhibits over-retention and/or over-promotion. For example, suppose we require HR policy to not make a loss with respect to any type of worker by adding the following set of constraints to the firm's original problem:

$$\text{supp}\left(\varphi\left(\cdot|\theta\right)\right) \in \underset{j\in\{0,1,2\}}{\arg\max}\ v_j\left(\theta\right) - c\left(\mathbb{E}_{\varphi}\left[\tilde{\theta}|j\right]\right)\ \forall \theta \in \Theta.$$

These incentive compatibility constraints that are standard in signalling games remove the possibility of over-retention and over-promotion because doing so necessarily results in the firm making losses on the types of workers it over-retains or over-promotes. In fact, without the ability to commit, we obtain the classic result as Waldman (1984) does that the firm always under-promotes (and also that the firm under-retains).

To see why, recall Figure 2 which shows the case in which the firm slightly over-retains by choosing $\theta_0 > \overline{\theta}_0$ and makes on some workers represented by the lower shaded triangle. In fact, the firm can eliminate losses by choosing a higher retention threshold (i.e., under-retaining further) so that the outside option of the expected type of a junior worker, $c(\mathbb{E}[\theta|\theta_0 \leq \theta < \theta_1])$, exactly equals the value of the marginally retained junior worker, $v_1(\theta_0)$.

5 Discussions

5.1 Job design

An alternative way to attenuate the positive signalling effect of promotion is to combine the jobs associated with positions 1 and 2 into one role, say to position $j = m$ (for *merged* position) as such merging of positions prevents the labour market from distinguishing workers

based on whether they are retained as a junior or promoted to be a senior. However, designing jobs in this manner would presumably be inefficient for the firm,[37] and we capture this inefficiency by assuming that a type-θ worker's output in the merged position is a weighted average of v_1 and v_2; i.e.,

$$v_m(\theta) = \alpha v_1(\theta) + (1-\alpha) v_2(\theta),$$

where $\alpha \in (0,1)$.[38] However, the firm cannot benefit from merging jobs in this manner. To see this, define $V_m(E) \coloneqq [v_m(E) - c(E)]_+$ as the firm's value function G^* denote a solution to the firm's problem when merging the two positions. But because $V_m \leq V$, we have

$$\begin{aligned}\int_\Theta V_m(E)\,\mathrm{d}G^*(E) &\leq \int_\Theta V(E)\,\mathrm{d}G^*(E) \\ &\leq \max_{G\in \mathrm{MPC}(F)} \int_\Theta V(E)\,\mathrm{d}G(E) = \widehat{V}^*.\end{aligned}$$

Hence, the firm cannot do better by merging positions. Thus, while the firm benefits from obfuscating workers' productivity to the labour market by misallocating across positions, it prefers to specialise workers into positions.

5.2 Vacuous job titles

In the National Longitudinal Survey of Youth in 1990, Pergamit and Veum (1999) finds that approximately a third of those who are promoted "continued to perform basically the same duties as before." [39] While we would interpret such a finding as suggestive of the fact that firms do indeed over-promote workers, it is also possible that the finding represents evidence of a disconnect in the relationship between job titles and job responsibilities.

In the information design game that we used to solve the firm's original problem, the firm can provide more information using messages ($m \in M$) than possible via the assignment of positions. However, we show that the firm has no incentive to do so; in our model, the firm sends at most one message for each position.

What drives this result is the fact that our set of assumptions implies that the firm's

[37] We note that the firm benefits from the ability to secretly assign positions $j \in \{1,2\}$ to attenuate the signalling effect without suffering a cost in terms of worker output.

[38] The results also hold when α depends continuously on θ.

[39] Somewhat more recently, Kosteas (2011) reports a similar finding based on the 1996–2006 waves of the National Longitudinal Surveys of Youth 1979 cohort.

value function, V, is not strictly convex on any nondegenerate interval of Θ.[40] If we were to make the alternative assumption that the set of production functions associated with outside firms is finite, then c and, consequently, V will be piecewise affine (but still convex). Consequently, in case the firm finds it optimal to induce posterior means in an affine region of V, it would be indifferent between not revealing and (partially or) fully revealing types of workers in the affine region. To the extent that there are other rationales for job titles (e.g., workers value job titles per se), the firm may give different job titles (i.e., send different m's) to workers who have the same responsibilities (i.e., assigned the same j).

5.3 Labour market that values the workers more

In our main setup, we assumed that it is efficient for the current employer of the firm to retain the most productive workers; i.e., $v_2(1) \geq c(1)$. However, it is also possible that high-θ workers should be employed by other firms that value them more; i.e., $v_2(1) < c(1)$. In this case, define $\overline{\theta}_2 := \max\{\theta \in \Theta : v_2(\theta) = c(\theta)\}$ as the efficient threshold for letting go of workers rather than retaining them in position $j = 2$. Then, when the prior mean about workers' types is sufficiently high, it can now be optimal for the firm to adopt a deterministic HR policy characterised by a retention threshold $\theta_2^* \geq \overline{\theta}_2$ in which all workers of type $\theta < \theta_2^*$ are retained and assigned to position $j = 2$ and all workers of type $\theta \geq \theta_2^*$ are let go.

5.4 Beyond two jobs

Our approach to solving for the optimal HR policies extends easily to cases in which the firm can assign workers to more positions. We note that, for additional positions (interpreted as more senior positions in which workers are potentially more roductive than existing ones) to matter, they must have a lower intercept and a greater slope that existing jobs.[41] However, the tradeoffs involved between any two consecutive positions would resemble that between $j = 1$ and 2 as we describe above. Moreover, additional care must be taken to determine whether the firm would find it optimal to assign workers to each position.

[40] Recall that V is a continuous function that is constant and equals zero on $[0, \overline{\theta}_0]$, positive and strictly concave on $[\overline{\theta}_0, \overline{\theta}_1]$ and $[\overline{\theta}_1, 1]$ with a convex kink at $\overline{\theta}_1$.

[41] As Costa (1988) notes, otherwise, the firm will never assign workers to the additional jobs.

5.5 Counteroffers

To simplify exposition and to be consistent with Waldman (1984), we do not explicitly model the wage-setting game and assume that the firm commits to offering a wage to each worker so as to ensure that the worker cannot obtain a strictly higher wage from other firms in the labour market. An alternative interpretation of this assumption is that the firm is always willing to make a counteroffer up to the average output of workers in position j, $v_j(\mathbb{E}_\varphi[\theta|j])$ if an outside firm were to make a poaching offer of say w to a worker in position $j \in \{1,2\}$. This assumption is reasonable in the information design game as the firm learns no more information about the worker than the market. However, to the extent that the firm learns the workers' types before assigning positions to them, the firm may be tempted to let go of workers who are worth less than w to the firm; i.e., let go of workers in position j whose types θ is such that $v_j(\theta) \leq w$. This type of counteroffer behaviour by the firm introduces an adverse selection problem for those attempting to poach workers from the firm (Greenwald, 1986; Gibbons and Katz, 1991). While a detailed analysis of the model with adverse selection is beyond the scope of this paper, we note that the firm benefits from letting go of workers strategically because doing so lowers the poaching offers the labour market is willing to make for the workers.

5.6 Relation with Waldman (1984) and Gibbons and Katz (1991)

Waldman (1984) studies how the firm's decision to promote or not promote a worker affects the wage it must pay in order to retain workers. In Waldman (1984), the firm assigns job $j \in \{1,2\}$ to the worker, where the firm's direct payoff from assigning job j to a type-θ worker is given by $(s_1,k_1) = (x(1+s),0)$ and $(s_2,k_2) = (0,1+s)$ for some $s > 0$ and $x \in (\mathbb{E}[\theta],1]$. The labour market's value for a type-θ worker is simply $c(\theta) \coloneqq \theta$. Thus, the firm's value function in Waldman (1984) is given by

$$V(E) = (1+s)\max\{E,x\} - E = \begin{cases} (1+s)x - \theta & \text{if } \theta < \overline{\theta}_1 \\ s\theta & \text{if } \theta \geq \overline{\theta}_1 \end{cases}, \tag{16}$$

where $\overline{\theta}_1 = x$. Observe that $V(\cdot) > 0$ so that the efficient outcome is for the incumbent firm to retain all workers. Because V is convex, a full revelation is optimal for the firm. However, because V is piecewise linear, the firm can attain the optimal payoff by simply setting the promotion threshold to be $\overline{\theta}_1$. Thus, the optimal promotion rule in Waldman

(1984) is efficient and does not exhibit under- or over-promotion.

Gibbons and Katz (1991) study a model that can capture both the signalling effect of retention à la Waldman (1984) and the adverse-selection effect of retention à la Greenwald (1986). The firm decides whether to retain or let go of the worker. The firm's payoffs from assigning each position are characterised, respectively, by $(s_0, k_0) = (0, 0)$ and $(s_1, k_1) = (s, 1)$, where $0 < s < \mathbb{E}[\theta]$. The labour market's value for a type-θ worker is again $c(\theta) \coloneqq \theta$. Hence, the firm's value function is simply $V(E) = s$. Observe that it is always efficient for the firm to retain workers of all types.

Observe that both Waldman (1984) and Gibbons and Katz (1991) assume a common production technology for outside firms. The analysis here, therefore, highlights that when the firm can commit to HR policies, inefficiencies in HR policies arise from the fact that heterogeneity in the production technologies of outside firms, which, in turn, leads to convexity in the workers' outside options.

6 Conclusion

In this paper, we show that an employer can find it optimal to implement a retention and promotion policy that over-retains and possibly over-promotes workers because doing so can reduce the wage costs for the firm by dampening the positive signalling effects of retention and promotions. Our results can provide a rationale for why firms may over-retain and over-promote workers as predicted by the Peter Principle. They can also be thought of as having a normative implication for firms when they design career paths for workers while taking into account changes in the workers' outside options from retention and promotion decisions. The results can also be viewed as a benchmark to assess alternative HR policies. This paper also highlights a gap in the empirical literature: the need to test for—not only the presence as the literature has done to date—but also of the direction of inefficiency associated with signalling effects of promotions and retentions.

While we interpret our model in the context of an incumbent employer's incentive to retain and promote its workers, the model can also be applied to other settings such as insurance to study incumbent insurer's incentive to under- and over-insure its insurees. Finally, we note that our method of solving signalling games in which the sender is able to commit to signalling strategies can be applied more generally to other signalling games.

References

Arieli, Itai, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita. 2023. "Optimal persuasion via bi-pooling." *Theoretical Economics*, 18(1): 15–36.

Baker, George P., Michael C. Jensen, and Kevin J. Murphy. 1988. "Compensation and Incentives: Practice vs. Theory." *The Journal of Finance*, 43(3): 593–616.

Benson, Alan, Danielle Li, and Kelly Shue. 2019. "Promotions and the Peter Principle." *The Quarterly Journal of Economics*, 134(4): 2085–2134.

Bernhardt, Dan. 1995. "Strategic Promotion and Compensation." *The Review of Economic Studies*, 62(2): 315–339.

Best, James, and Daniel Quigley. 2024. "Persuasion for the Long Run." *Journal of Political Economy*, 132(5): 1740–1791.

Blume, Andreas. 2003. "Bertrand without fudge." *Economics Letters*, 78(2): 167–168.

Bognanno, Michael, and Eduardo Melero. 2016. "Promotion Signals, Experience, and Education." *Journal of Economics & Management Strategy*, 25(1): 111–132.

Cassidy, Hugh, Jed DeVaro, and Antti Kauhanen. 2016. "Promotion signaling, gender, and turnover: New theory and evidence." *Journal of Economic Behavior & Organization*, 126: 140–166.

Costa, Joan E. Ricart I. 1988. "Managerial Task Assignment and Promotions." *Econometrica*, 56(2): 449–466.

Dato, Simon, Andreas Grunewald, Matthias KrÃ€kel, and Daniel MÃŒller. 2016. "Asymmetric employer information, promotions, and the wage policy of firms." *Games and Economic Behavior*, 100: 273–300.

Deb, Rahul, Mallesh Pai, and Maher Said. 2023. "Indirect Persuasion." *Working Paper*.

DeVaro, Jed, and Michael Waldman. 2012. "The Signaling Role of Promotion: Further Theory and Empirical Evidence." *Journal of Labor Economics*, 30(1): 91–147.

Dworczak, Piotr, and Giorgio Martini. 2019. "The Simple Economics of Optimal Persuasion." *Journal of Political Economy*, 127(5): 1993–2048.

Ghosh, Suman, and Michael Waldman. 2010. "Standard promotion practices versus up-or-out contracts." *The RAND Journal of Economics*, 41(2): 301–325.

Gibbons, Robert, and John Roberts, ed. 2013. *The Handbook of Organizational Economics.* . 1 ed., Princeton University Press.

Gibbons, Robert, and Lawrence F. Katz. 1991. "Layoffs and Lemons." *Journal of Labor Economics*, 9(4): 351–380.

Gibbons, Robert, and Michael Waldman. 2006. "Enriching a Theory of Wage and Promotion Dynamics inside Firms." *Journal of Labor Economics*, 24(1): 59–107.

Goldstein, Itay, and Yaron Leitner. 2018. "Stress tests and information disclosure." *Journal of Economic Theory*, 177: 34–69.

Greenwald, Bruce C. 1986. "Adverse Selection in the Labour Market." *The Review of Economic Studies*, 53(3): 325–347.

Ishida, Junichiro. 2004. "Signaling and strategically delayed promotion." *Labour Economics*, 11(6): 687–700.

Kahn, Lisa B. 2013. "Asymmetric Information between Employers." *American Economic Journal: Applied Economics*, 5(4): 165–205.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review*, 101(6): 2590–2615.

Kosteas, Vasilios D. 2011. "Job Satisfaction and Promotions." *Industrial Relations: A Journal of Economy and Society*, 50(1): 174–194.

Lazear, Edward P. 2004. "The Peter Principle: A Theory of Decline." *Journal of Political Economy*, 112(S1): S141–S163.

Lin, Xiao, and Ce Liu. 2024. "Credible Persuasion." *Journal of Political Economy*, 132(7): 2228–2273.

Mathevet, Laurent, David Pearce, and Ennio Stacchetti. 2022. "Reputation for a Degree of Honesty." *Working Paper*.

Mensch, Jeffrey. 2021. "Monotone persuasion." *Games and Economic Behavior*, 130: 521–542.

Milgrom, Paul, and John Roberts. 1988. "An Economic Approach to Influence Activities in Organizations." *American Journal of Sociology*, 94: S154–S179.

Milgrom, Paul, and Sharon Oster. 1987. "Job Discrimination, Market Forces , and the Invisibility Hypothesis." *The Quarterly Journal of Economics*, 102(3): 453.

Owan, Hideo. 2004. "Promotion, Turnover, Earnings, and Firm-Sponsored Training." *Journal of Labor Economics*, 22(4): 955–978.

Pergamit, Michael R., and Jonathan R. Veum. 1999. "What is a Promotion?" *Industrial and Labor Relations Review*, 52(4): 581.

Peter, Laurence J., and Raymond Hull. 1969. *The Peter Principle.* New York:William Morrow.

Pinkston, Joshua C. 2009. "A Model of Asymmetric Employer Learning with Testable Implications." *Review of Economic Studies*, 76(1): 367–394.

Schonberg, Uta. 2007. "Testing for Asymmetric Employer Learning." *Journal of Labor Economics*, 25: 651–691.

Shaked, Moshe, and George J. Shanthikumar, ed. 2007. "Stochastic Orders." , ed. Moshe Shaked and George J. Shanthikumar . 1 ed. Springer New York, NY.

Vohra, Akhil. 2024. "Unraveling and Inefficient Matching." *American Economic Journal: Microeconomics (forthcoming)*.

Waldman, Michael. 1984. "Job Assignments, Signalling, and Efficiency." *The RAND Journal of Economics*, 15(2): 255.

Waldman, Michael. 2003. "Ex Ante versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion." *Economic Inquiry*, 41(1): 27–41.

A Appendix

A.1 Proof of Lemma 1

We first summarise results from Dworczak and Martini (2019) and Arieli et al. (2023) that are relevant in our proofs for convenience. The following set of facts follows directly from theorems 1 and 2 of Dworczak and Martini (2019).

Fact 1 (Dworczak and Martini, 2019)**.** *A CDF G is a solution to the firm's information design problem,* (3)*, if there exists a function $p : \Theta \to \mathbb{R}$ such that*

$$p \geq V \text{ and } p \text{ is convex}, \tag{17}$$

$$\operatorname{supp}(G) \subseteq \{E \in \Theta : V(E) = p(E)\}, \tag{18}$$

$$\int_\Theta p(E)\,\mathrm{d}G(E) = \int_\Theta p(\theta)\,\mathrm{d}F(\theta), \tag{19}$$

$$G \in \mathrm{MPC}(F). \tag{20}$$

Moreover, if V is Lipschitz continuous, then the converse also holds; i.e., if a CDF G solves (3)*, then there exists a function $p : \Theta \to \mathbb{R}$ that satisfies* (17)–(20)*.*

Given a pair (G, p) that satisfies (17)–(20), we refer to p as a *support function of* G.[42]

Following Arieli et al. (2023), we call a mean-preserving contraction G of F, $G \in \mathrm{MPC}(F)$, a *bi-pooling* distribution (with respect to F) if G can be induced as a distribution of posterior means by an information structure that, for some collection (indexed by $\mathscr{I}$) of pairwise disjoint open intervals $\bigcup_{I\in\mathscr{I}} I \subseteq \Theta$: (i) partitions Θ into $\bigcup_{I\in\mathscr{I}} I$ and its complement; (ii) whenever $\theta \in \bigcup_{I\in\mathscr{I}} I$, the information structure reveals which interval the state is in and provides an additional binary signal; and (iii) the sender fully reveals the state whenever the state is in the complement of $\bigcup_{I\in\mathscr{I}} I$. If a nontrivial binary signal is provided in an interval $I \in \mathscr{I}$, we call the interval I a *bi-pooling* interval; if only a trivial binary signal is provided, then we call I a *pooling* interval. We call an interval $I \in \mathscr{I}$ in which the sender fully reveals the state a *separating* interval. If all intervals are pooling intervals, we refer to G as a *pooling* distribution.

The following set of facts is a combination of Lemma 1, Theorem 1 and Proposition 3 in Arieli et al. (2023).

[42] Dworczak and Martini (2019) refers to p instead as a price function for reasons that are not relevant for our paper.

Fact 2 (Arieli et al., 2023)**.**

(i) A bi-pooling distribution G solves (3)*.*

(ii) Suppose V is upper semicontinuous and there exists a collection of $J \in \mathbb{N}$ points $\{\overline{\theta}_j\}_{j=-1}^{J}$, with $0 = \overline{\theta}_{-1} < \overline{\theta}_0 < \cdots < \overline{\theta}_J = 1$, such that V is either concave or convex in the interval $(\overline{\theta}_{j-1}, \overline{\theta}_j)$ for every $j \in \{0,1,\ldots,J\}$. Let $p : \Theta \to \mathbb{R}$ denote the support function of an optimal bi-pooling distribution G. Then, there exists a collection of functions, $\{p_j : [\theta_{j-1}, \theta_j] \to \mathbb{R}\}_{j=0}^{J}$, where $0 = \theta_{-1} < \theta_0 < \ldots < \theta_m = 1$ for some $m \leq 2(J+1)-1$,[43] such that $p|_{[\theta_{j-1},\theta_j]} = p_j$ for all $j \in \{0,1,\ldots,J\}$, and each p_j satisfies exactly one of the following conditions:

(a) $p_j = V|_{[\theta_{j-1},\theta_j]}$;

(b) $p_j \geq V|_{[\theta_{j-1},\theta_j]}$, and p_j is affine and tangential to V at $\mathbb{E}[\theta|\theta_{j-1} < \theta < \theta_j]$; or

(c) $p_j \geq V|_{[\theta_{j-1},\theta_j]}$, and p_j affine and tangential to V at two distinct points $\underline{E}_j$ and $\overline{E}_j$ such that

$$\theta_{j-1} \leq \underline{E}_j \leq \mathbb{E}\left[\theta|\theta_{j-1} < \theta < \theta_j\right] \leq \overline{E}_j \leq \mathbb{E}\left[\theta|\theta'_{j-1} < \theta < \theta_j\right] \leq \theta_j, \quad (21)$$

where θ'_{j-1} is defined implicitly via $\mathbb{E}[\theta|\theta_{j-1} < \theta < \theta'_{j-1}] = \underline{E}_j$.

For each of the cases (a) to (c), we call the interval $[\theta_{j-1}, \theta_j]$ a *separating*, *pooling* and *bi-pooling* interval, respectively. Moreover, in case $[\theta_{j-1}, \theta_j]$ is a bi-pooling interval, we refer to p_j (extended to $\mathbb{R}$) as the *bi-tangent* of V on the interval $[\theta_{j-1}, \theta_j]$, and the points $\underline{E}_j$ and $\overline{E}_j$ as the bi-tangency points.

Recall that a tuple $(\theta_0, \theta_1, E_2) \in \Theta^3$ is feasible if $\theta_0 \in [0, \overline{\theta}_0]$, $\theta_1 \in [\theta_0, 1]$, and $\mathbb{E}[\theta|\theta > \theta_0] \leq E_2 \leq \mathbb{E}[\theta|\theta \geq \theta_1]$, and a CDF G is feasible if there exists a feasible tuple $(\theta_0, \theta_1, E_2)$ such that $G = G_{\theta_0,\theta_1,E_2}$, where $G_{\theta_0,\theta_1,E_2}$ is a CDF such that (i) $G \in \text{MPC}(F)$; (ii) the mass on $[0, \theta_0)$ from F is put on $\mathbb{E}[\theta|\theta < \theta_0]$; and (iii) the mass on $[\theta_0, 1]$ from F is put on $\mathbb{E}[\theta|\theta_0 < \theta < \theta_1]$ and on E_2. Observe that, for any feasible tuple $(\theta_0, \theta_1, E_2)$, the CDF $G_{\theta_0,\theta_1,E_2}$ is a bi-pooling distribution in which $[0, \theta_0)$ is a pooling interval and

- if $E_2 = \mathbb{E}[\theta|\theta > \theta_0]$ (in which case $\theta_0 = \theta_1$), $[\theta_0, 1]$ is a pooling interval;
- if $E_2 = \mathbb{E}[\theta|\theta \geq \theta_1]$, $[\theta_0, \theta_1)$ and $[\theta_1, 1]$ are pooling intervals;

[43] While Proposition 3 in Arieli et al. (2023) states $m \leq 2(J+1)$ instead, the argument in their proof in fact implies $m \leq 2(J+1)-1$.

- otherwise $[\theta_0, 1]$ is a bi-pooling interval because $\mathbb{E}[\theta|\theta_0 < \theta < \theta_1]$ and E_2 are bi-tagency points.[44]

The observation above means that feasible CDFs, which are parameterised by a family of feasible tuples $(\theta_0, \theta_1, E_2)$, is a class of bi-pooling distributions.

We first establish the necessary and sufficient conditions for V to have a bi-tangent on $[\overline{\theta}_0, 1]$. We note that, because V is constant and equals zero on $[0, \overline{\theta}_0]$, V always has a bi-tangent on any subinterval of $[0, \overline{\theta}_0]$.

Lemma 5. *V has a bi-tangent on $[\overline{\theta}_0, 1]$ if and only if (i) the tangent to V_1 at $\overline{\theta}_0$ lies above V_2 on $[\overline{\theta}_0, 1]$ and (ii) the tangent to V_2 at 1 lies above V_1 on $[\overline{\theta}_0, 1]$. Moreover, when it exists, the bi-tangent is unique, and it is tangent to V at some $E_1 \in [\overline{\theta}_0, \overline{\theta}_1)$ and $E_2 \in [\overline{\theta}_1, 1]$.*

Proof. Suppose a bi-tangent on $[\overline{\theta}_0, 1]$ exists; i.e., there exist $E_1, E_2 \in [\overline{\theta}_0, 1]$ with $E_2 > E_1$ such that the tangents to V at E_1 and E_2 coincide. Recall that

$$V(\theta) = \begin{cases} V_1(\theta) & \text{if } \theta \in \left[\overline{\theta}_0, \overline{\theta}_1\right], \\ V_2(\theta) & \text{if } \theta \in \left[\overline{\theta}_1, 1\right], \end{cases}$$

and that V_1 and V_2 are strictly concave. It cannot be that $E_1, E_2 \in [\overline{\theta}_0, \overline{\theta}_1]$: strict concavity means that each point on V_j has a unique tangent so that tangents to V_1 at two distinct points E_1 and E_2 cannot coincide. Similarly, it cannot be that $E_1, E_2 \in [\overline{\theta}_1, 1]$. Thus, we must have $E_1 \in [\overline{\theta}_0, \overline{\theta}_1]$ and $E_2 \in [\overline{\theta}_1, 1]$. For each $j \in \{1, 2\}$ and $E \in \Theta$, let $t_j(\cdot|E) : \Theta \to \mathbb{R}$ denote the tangent to V_j at point E; i.e.,

$$t_j(\theta|E) \coloneqq V_j'(E)(\theta - E) + V_j(E).$$

Since E_1 and E_2 are bi-tangency points, we must have $\ell \coloneqq V_1'(E_1) = V_2'(E_2)$ and $t_1(\cdot|E_1) = t_2(\cdot|E_2)$. Because V_1 is strictly concave and $\overline{\theta}_0 \leq E_1$,

$$t_1\left(\theta|\overline{\theta}_0\right) \geq t_1(\theta|E_1) = t_2(\theta|E_2) \ \forall \theta \geq E_1.$$

[44] To see this, observe that:

$$\theta_0 \leq \mathbb{E}[\theta|\theta_0 < \theta < \theta_1] \leq \underbrace{\mathbb{E}[\theta|\theta_0 < \theta < 1]}_{=\mathbb{E}[\theta|\theta > \theta_0]} \leq E_2 \leq \underbrace{\mathbb{E}[\theta|\theta_1 < \theta < 1]}_{=\mathbb{E}[\theta|\theta \geq \theta_1]} < 1.$$

Because V_2 is strictly concave, V_2 must lie below the tangent line $t_1(\cdot|\overline{\theta}_0)$. Given that $V_2(\theta) \leq V_1(\theta)$ for any $\theta < \overline{\theta}_0$, it follows that $t_1(\cdot|\overline{\theta}_0)$ lies above V_2 on $[\overline{\theta}_0, 1]$. Analogously, because V_2 is strictly concave and $1 \geq E_2$,

$$t_2(\cdot|1) \geq t_2(\theta|E_2) = t_1(\theta|E_1) \ \forall \theta \leq E_2.$$

Because V_1 is strictly concave, V_1 must lie below the tangent line $t_2(\cdot|1)$. Given that $V_2(\theta) \geq V_1(\theta)$ for any $\theta \geq \overline{\theta}_1$, it follows that $t_2(\cdot|1)$ lies above V_1 on $[\overline{\theta}_0, 1]$.

To prove the converse, let $t_1(\cdot)$ and $t_2(\cdot)$ denote the tangent to V_1 at $\overline{\theta}_0$ and tangent to V_2 at 1, respectively. Suppose that t_j lies above $V_{j'}$ on $[\overline{\theta}_0, 1]$ for each distinct $j, j' \in \{1,2\}$. Then, $V_1'(\theta_1) > V_2'(1)$. By definition of $\overline{\theta}_1$, $V_1'(\overline{\theta}_1) < V_2'(\overline{\theta}_2)$. By strict concavity of V_1 and V_2, there must exist a unique bi-tangent. ■

Recall $c_E(\cdot)$ denotes the tangent to c at E. Then, condition (i) of Lemma 5 holds if and only if

$$c(\theta) - c_{\overline{\theta}_0}(\theta) \geq v_2(\theta) - v_1(\theta) \ \forall \theta \in \left[\overline{\theta}_1, 1\right], \tag{22}$$

where we note that $v_2(\cdot) - v_1(\cdot) \geq 0$ on $[\overline{\theta}_1, 1]$. Condition (ii) of Lemma 5 holds if and only if

$$v_1(\theta) - v_2(\theta) \geq c_1(\theta) - c(\theta) \ \forall \theta \in \left[\overline{\theta}_0, \overline{\theta}_1\right], \tag{23}$$

where we note that $v_1(\cdot) - v_2(\cdot) \geq 0$ on $[\overline{\theta}_0, \overline{\theta}_1]$. The first condition requires c to exhibit sufficient convexity, while the second condition limits how much convexity c can exhibit. Thus, a bi-tangent exists whenever c is sufficiently convex but not too convex; in other words, a bi-tangent exists whenever the increase in the marginal product of labour is not too little or too much.

We are now ready to prove Lemma 1.

Lemma 1. *There exists a feasible G^* that solves the information design problem,* (3)*; i.e.,*

$$\widehat{V}^* = \max_{(\theta_0, \theta_1, E_2)} \int_\Theta V(E) \, \mathrm{d}G_{\theta_0, \theta_1, E_2}(\theta) \ \textit{s.t.} \ (\theta_0, \theta_1, E_2) \ \textit{is feasible}. \tag{24}$$

Proof. By Fact 2(i), a bi-pooling G solves (3). Because V is continuous, linear on $[\overline{\theta}_{-1} := 0, \overline{\theta}_0]$, strictly convex on the intervals $[\overline{\theta}_0, \overline{\theta}_1]$ and $[\overline{\theta}_1, \overline{\theta}_2 := 1]$, and has convex kinks at $\overline{\theta}_0$ and $\overline{\theta}_1$, by Fact 2(ii), any bi-pooling interval can be found by identifying the bi-tangents of V.

Consider first the case in which V has a bi-tangent on $[\overline{\theta}_0, 1]$. By Lemma 5, the bi-tangent is unique and we denote it as $V^b(\cdot)$. Let θ_0 be such that $V^b(\theta_0) = 0$. Given the assumptions on V, $\theta_0^b < \overline{\theta}_0$. Denote the two bi-tangency points as $E_1 \in [\overline{\theta}_0, \overline{\theta}_1)$ and $E_2 \in [\overline{\theta}_1, 1]$. If E_1 and E_2 satisfy (21), then $[\theta_0^b, 1]$ is a bi-pooling interval. Since V is constant and equals zero on $[0, \overline{\theta}_0]$, $\max\{V^b(\cdot), 0\}$ is function satisfying (17)–(20). Moreover, G that can be supported by such a p is associated with $(\theta_0, \theta_1, E_2)$, where θ_1 satisfies $\mathbb{E}[\theta|\theta_0 < \theta < \theta_1] = E_2$; i.e., G is feasible.

If an optimal G is not bi-pooling (either because a bi-tangent does not exist or even if it exists, the tangency points, E_1 and E_2, do not satisfy (21)), then G must be a pooling distribution. By Fact 2(ii), a support function of an optimal distribution G, $p : \Theta \to \mathbb{R}$, is a concatenation of at most 5 affine functions; i.e., there exists a collection $\{p_j : [\theta_{j-1}, \theta_j] \to \mathbb{R}\}_{j=0}^4$, where

$$0 = \theta_{-1} < \theta_0 < \overline{\theta}_0 < \theta_1 < \theta_2 < \overline{\theta}_1 < \theta_3 < \theta_4 = 1, \tag{25}$$

such that $p|_{[\theta_{j-1}, \theta_j]} = p_j$ for each $j \in \{0, 1, \ldots, 4\}$. Moreover, each p_j is such that either $p_j = V|_{[\theta_{j-1}, \theta_j]}$ and p_j is convex (i.e., $[\theta_{j-1}, \theta_j]$ is a separating interval), or $p_j \geq V_{|_{[\theta_{j-1}, \theta_j]}}$ and p_j is affine and tangent to V at $\mathbb{E}[\theta|\theta_{j-1} \leq \theta \leq \theta_j]$ (i.e., $[\theta_{j-1}, \theta_j]$ is a pooling interval). Since V is convex (in fact, linear) on $[0, \overline{\theta}_0]$, p_0 must be a constant function equal to zero. Furthermore, to ensure that p is convex and $p \geq V$, p_1 must be tangent to V at some $E_1 \in [\overline{\theta}_0, \overline{\theta}_1]$, and p_4 must be tangent to V at some $E_2 \in [\overline{\theta}_1, 1]$. However, because V is strictly concave on $[\overline{\theta}_0, \overline{\theta}_1]$ and on $[\overline{\theta}_1, 1]$, and has a convex kink at $\overline{\theta}_1$, p_2, if it exists, would lie strictly above V; contradicting that p_2 is tangential to V on $[\theta_1, \theta_2]$. Moreover, any point in which p_3 is tangential to V would be a concave kink; contradicting the convexity of p. Thus, it follows that p must be a concatenation of at most three affine functions. Moreover, any G that is pooling that can be supported by such a p is associated with $(\theta_0, \theta_1, \mathbb{E}[\theta|\theta \geq \theta_1])$ and so feasible. ■

A.2 Proof of Lemma 2

Lemma 2. *The value of the firm's problem,* (2)*, in the original game equals the value of the firm's problem,* (3)*, in the related game; i.e., $V^* = \widehat{V}^*$.* Moreover, if G^* *solves* (3)*, then there exists an HR policy $\varphi^* \in \Psi$ that solves the firm's original problem,* (2)*, and induces G^* as the distribution of posterior means.*

Proof. We first show that $\widehat{V}^* \leq V^*$. Suppose $\widehat{G}^* \in \text{MPC}(F)$ solves the firm's information

design problem, (3). Define

$$\Theta_0 := \left\{ E \in \operatorname{supp}\left(\widehat{G}^*\right) : \max_{j\in\{1,2\}} v_j(E) \leq c(E) \right\},$$
$$\Theta_1 := \left\{ E \in \operatorname{supp}\left(\widehat{G}^*\right) \setminus \Theta_0 : v_2(E) \leq v_1(E) \right\},$$
$$\Theta_2 := \left\{ E \in \operatorname{supp}\left(\widehat{G}^*\right) \setminus \Theta_0 : v_1(E) < v_2(E) \right\},$$

and let

$$E_j^* := \frac{\int_{\Theta_j} E \mathrm{d}\widehat{G}^*(E)}{\widehat{G}^*\left(\Theta_j\right)} \ \forall j \in \{0,1,2\}.$$

By our choice of E_j^*'s, there exists $G^* \in \mathrm{MPC}(F)$ so that $\operatorname{supp}(G^*) = \{E_0^*, E_1^*, E_2^*\}$. Then, by construction, and the fact that v is affine and c is convex, we have

$$\widehat{V}^* \leq \int_\Theta V(E)\,\mathrm{d}G^*(E) = \sum_{j=1}^{2} \left[v_j\left(E_j^*\right) - c\left(E_j^*\right) \right] G\left(\left\{E_j^*\right\}\right).$$

Let $\pi^* \in \Pi$ denote an experiment with $\operatorname{supp}(\pi^*) = \{m_0, m_1, m_2\}$ such that

$$\mathbb{E}_{\pi^*}\left[\theta | m_j\right] = E_j^* \ \forall j \in \{0,1,2\},$$

which exists because $G^* \in \mathrm{MPC}(F)$. Then, $\int_\Theta \pi^*(m_j|\theta)\mathrm{d}F(\theta) = G^*(\{E_j^*\})$ and so

$$\begin{aligned}
\widehat{V}^* &\leq \sum_{j=1}^{2} \left[v_j\left(E_j^*\right) - c\left(E_j^*\right) \right] \int_\Theta \pi^*\left(m_j|\theta\right) \mathrm{d}F(\theta) \\
&= \sum_{j=1}^{2} \left[s_j + k_j \frac{\int_\Theta \theta \pi^*\left(m_j|\theta\right) \mathrm{d}F(\theta)}{\int_\Theta \pi^*\left(m_j|\theta\right) \mathrm{d}F(\theta)} - c\left(E_j^*\right) \right] \int_\Theta \pi^*\left(m_j|\theta\right) \mathrm{d}F(\theta) \\
&= \int_\Theta \sum_{j=1}^{2} \left[\underbrace{s_j + k_j\theta}_{=v_j(\theta)} - c\left(E_j^*\right) \right] \pi^*\left(m_j|\theta\right) \mathrm{d}F(\theta) \\
&\leq \max_{\varphi:\Theta\to\Delta\{0,1,2\}} \int_\Theta \sum_{j=1}^{2} \left[v_j(\theta) - c\left(\mathbb{E}_\varphi\left[\theta|j\right]\right) \right] \varphi(j|\theta)\,\mathrm{d}F(\theta) = V^*,
\end{aligned}$$

where the last inequality follows because we can choose $\varphi^* : \Theta \to \Delta\{0,1,2\}$ as

$$\varphi^*(j|\cdot) = \pi^*\left(m_j|\cdot\right) \ \forall j \in \{0,1,2\} \tag{26}$$

so that $\mathbb{E}_{\varphi^*}[\theta|j] = E_j^*$.

We now show that $V^* \leq \widehat{V}^*$. Let φ^* solve the firm's original problem, (2), and denote $E_j^* \coloneqq \mathbb{E}_{\varphi^*}[\theta|j]$. Let $\{m_0, m_1, m_2\}$ be a set of three distinct elements of M and construct a Blackwell experiment $\pi^* \in \Pi$ via (26). Let us first show that $v_j(E_j^*) - c(E_j^*) \geq 0$ for $j \in \{1,2\}$. By way of contradiction, suppose for $\widehat{j} \in \{1,2\}$, $v_{\widehat{j}}(E_{\widehat{j}}^*) - c(E_{\widehat{j}}^*) < 0$. Consider $\widehat{\varphi} : \theta \to \mathbb{R}$ such that

$$\widehat{\varphi}\left(j'|\cdot\right) \coloneqq \begin{cases} \varphi^*\left(0|\cdot\right) + \varphi^*\left(\widehat{j}|\cdot\right) & \text{if } j' = 0, \\ 0 & \text{if } j' = \widehat{j}, \\ \varphi^*\left(j'|\cdot\right) & \text{otherwise.} \end{cases}$$

Letting $\widehat{j}' \in \{1,2\}$ with $\widehat{j}' \neq \widehat{j}$, we have

$$\begin{aligned} V\left(\widehat{\varphi}\right) - V\left(\varphi^*\right) &= \int_\theta \sum_{j=1}^{2} \left[v_j\left(\theta\right) - c\left(E_j^*\right)\right] \left[\widehat{\varphi}\left(j|\theta\right) - \varphi^*\left(j|\theta\right)\right] \mathrm{d}F\left(\theta\right) \\ &= -\int_\Theta \left[v_{\widehat{j}}(\theta) - c\left(E_{\widehat{j}}^*\right)\right] \varphi^*\left(\widehat{j}|\theta\right) \mathrm{d}F\left(\theta\right) > 0, \end{aligned}$$

which contradicts the optimality of φ^*. Letting $\pi^*(m_j) \coloneqq \int_\Theta \pi^*(m_j|\theta)\mathrm{d}F(\theta)$, we have

$$\begin{aligned} V^* &= \int_\Theta \sum_{j=1}^{2} \left[v_j\left(\theta\right) - c\left(E_j^*\right)\right] \varphi^*\left(j|\theta\right) \mathrm{d}F\left(\theta\right) \\ &= \sum_{j=1,2} \left[v_j\left(E_j^*\right) - c\left(E_j^*\right)\right] \pi^*\left(m_j\right) \\ &\leq \sum_{j=1,2} \left[\max_{j' \in \{1,2\}} v_{j'}\left(E_j^*\right) - c\left(E_j^*\right)\right] \pi^*\left(m_j\right) \\ &\leq \max_{G \in \mathrm{MPC}(F)} \int_\Theta V\left(E\right) \mathrm{d}G\left(E\right) = \widehat{V}^*, \end{aligned}$$

where the second follows from the fact that π^* induces $G^* \in \mathrm{MPC}(F)$ such that $\mathrm{supp}(G^*) = \{E_0^*, E_1^*, E_2^*\}$. ■

A.3 Proof of Lemma 3

Lemma 3. *Any feasible distribution of posterior means can be induced by a monotone HR policy.*

Proof. Let G denote a feasible CDF associated with a feasible tuple $(\theta_0, \theta_1, E_2)$. As we noted in Appendix A.1, G is a bi-pooling distribution. Thus, by lemma 2 in Arieli et al. (2023), G can be induced by a monotone information structure π. By construction, G induces at most three posterior means, E_0, E_1, E_2, such that $E_0 < E_1 \leq E_2$. Hence, it can be implemented by an HR policy φ by setting $\varphi(j|\theta) = \pi(m_j|\theta)$ for all $\theta \in \Theta$ and $j \in \{0,1,2\}$, where each m_j induces posterior mean E_j. ■

A.4 Proof of Proposition 1

Proposition 1. *A deterministic HR policy is optimal if and only if there does not exist a tuple $(\theta_0^b, E_1^b, E_2^b)$ with $\theta_0^b \leq E_1^b \leq E_2^b$ such that:*

(i) $c(E_j^b) - v_j(\theta_0^b) = c'(E_j^b)(E_j^b - \theta_0^b)$ *for* $j = \{1,2\}$*;*

(ii) $k_2 - k_1 = c'(E_2^b) - c'(E_1^b)$*; and*

(iii) $E_1^b \leq \mathbb{E}[\theta|\theta > \theta_0^b] < E_2^b < \mathbb{E}[\theta|\theta \geq \theta_1^b]$*, where* θ_1^b *satisfies* $\mathbb{E}[\theta|\theta_0^b < \theta < \theta_1^b] = E_1^b$.

Proof. We first note that any feasible G associated with $(\theta_0, \theta_1, E_2)$ that is a pooling distribution can be induced by a deterministic monotone HR policy that lets go of workers of types in $[0, \theta_0)$, retains in the junior position workers of types in $[\theta_0, \theta_1)$, and promotes to the senior position workers of types in $[\theta_1, 1]$. However, it follows from Lemma 3 in Arieli et al. (2023) (along with an argument analogous to the proof of Lemma 3 above) that G cannot be induced by any deterministic monotone HR policy if and only if G has a bi-pooling interval on $[\theta_0, 1]$. By Fact 2(ii), G has a bi-pooling interval on $[\theta_0, 1]$ if there exists two bi-tangecy points $E_1^b := \mathbb{E}[\theta|\theta_0 \leq \theta < \theta_1]$ and $E_2^b := E_2$ to V that satisfy

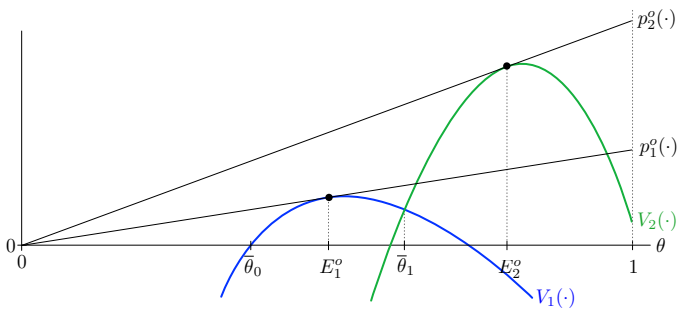
$$\theta \leq E_1^b \leq \mathbb{E}[\theta|\theta > \theta_0] \leq E_2^b \leq \mathbb{E}[\theta|\theta \geq \theta_1] \leq 1.$$

Letting $\theta_0^b := \theta_0$ gives condition (iii). For E_1^b and E_2^b to be bi-tangency points of V, it must be that $V'(E_1^b) = V'(E_2^b)$, which corresponds to condition (i), and that the bi-tangent to V equals zero at θ_0^b, which corresponds to condition (ii). ■

A.5 Optimal feasible G

As shown in Figure 5, for each $j \in \{1,2\}$, let $p_j^o : \Theta \to \mathbb{R}$ denote the smallest linear function that exceeds V_j, and let E_j^o denote the smallest $E \in \Theta$ such that $p_j^o(E) = V_j(E)$. Observe that while $p_2^o \geq p_1^o$ implies $\max V_2 > \max V_1$, the converse is not true.

Figure 5: p_j^o and E_j^o's.



We first consider the case when $p_2^o \geq p_1^o$.

Proposition 7. *Suppose $p_2^o \geq p_1^o$ and $\mathbb{E}[\theta] \geq E_2^o$. Then, $G = F$ solves the firm's problem and assigning all workers to position $j = 2$ is optimal.*

Proof. Because V_2 is concave on $[E_2^o, 1] \subseteq [\overline{\theta}_1, 1]$, then $G = F$ (i.e., providing no information) is optimal for the firm because it can be supported by p given by the tangent to V at $\mathbb{E}[\theta]$. Since $\mathbb{E}[\theta] \in [\overline{\theta}_1, 1]$, the firm maximises its expected payoff by assigning all workers to position $j = 2$. ■

Given a feasible G associated with $(\theta_0, \theta_1, E_2)$, we define the firm's expected payoff as:

$$\begin{aligned}\widehat{V}(\theta_0, \theta_1, E_2) &= \int_\Theta V(E)\, \mathrm{d}G_{\theta_0,\theta_1,E_2}(\theta) \\ &= (1 - F(\theta_0)) \left[V\left(\mathbb{E}[\theta | \theta_0 < \theta < \theta_1]\right) \alpha_{\theta_0,\theta_1,E_2} + V(E_2) \left(1 - \alpha_{\theta_0,\theta_1,E_1}\right) \right],\end{aligned}$$

where

$$\alpha_{\theta_0,\theta_1,E_2} := \begin{cases} 0 & \text{if } \theta_1 = 1 \text{ and } E_2 = \mathbb{E}[\theta | \theta > \theta_0], \\ \frac{E_2 - \mathbb{E}[\theta|\theta > \theta_0]}{E_2 - \mathbb{E}[\theta|\theta_0 < \theta < \theta_1]} & \text{otherwise.} \end{cases}$$

The problem for the firm now, (4), is to maximise $\widehat{V}$ with respect to $(\theta_0, \theta_1, E_2)$ among all feasible tuples. Since $\widehat{V}$ is differentiable, a necessary condition for optimality is that $(\theta_0, \theta_1, E_2)$ satisfies the first-order conditions (while accounting for boundary solutions). We establish sufficiency by constructing the support function for G associated with $(\theta_0, \theta_1, E_2)$ as an upper envelope of tangents to $\widehat{V}$ at $\mathbb{E}[\theta | \theta_0 < \theta < \theta_1]$ and E_2, while imposing an additional condition to ensure that the support function is convex. We say that a feasible tuple is *nondegenerate* if it is not associated with F.

Lemma 6. *A feasible nondegenerate tuple* $(\theta_0^*, \theta_1^*, E_2^*)$ *solves the firm's problem,* (4)*, if and only if it satisfies the first-order conditions and* $V'(\mathbb{E}[\theta|\theta \geq \theta_1^*]) \geq V'(\mathbb{E}[\theta|\theta_0^* < \theta < \theta_1^*])$.

Proof. The idea is to construct a supporting function p^* that supports $G_{(\theta_0^*, \theta_1^*, E_2^*)}$ using the first-order conditions to (4). Then, by Fact 2, it would follow that $(\theta_0^*, \theta_1^*, E_2^*)$ must be optimal. To that end, let us first fix $\theta_0 \in (0, \overline{\theta}_0)$ and $\theta_1 \in (\overline{\theta}_0, 1)$ and solve

$$\max_{E_2 \in \Theta} \widehat{V}(\theta_0, \theta_1, E_2) \text{ s.t. } \mathbb{E}[\theta|\theta > \theta_0] \leq E_2 \leq \mathbb{E}[\theta|\theta \geq \theta_1].$$

We denote the maximiser as $E_2^*(\theta_0, \theta_1)$. The derivative of the objective with respect to E_2 is

$$\frac{\mathrm{d}\widehat{V}(\cdot)}{\mathrm{d}E_2} = (1 - F(\theta_0))\left(1 - \alpha_{\theta_0, \theta_1, E_2}\right)\left[V'(E_2) - \frac{V(E_2) - V(\mathbb{E}[\theta|\theta_0 < \theta < \theta_1])}{E_2 - \mathbb{E}[\theta|\theta_0 < \theta < \theta_1]}\right].$$

If $\frac{\mathrm{d}\widehat{V}(\cdot)}{\mathrm{d}E_2}|_{E_2 = \mathbb{E}[\theta|\theta \geq \theta_1]} > 0$, we let $E_2^*(\theta_0, \theta_1) = \mathbb{E}[\theta|\theta \geq \theta_1]$ and if $\frac{\mathrm{d}\widehat{V}(\cdot)}{\mathrm{d}E_2}|_{E_2 = \mathbb{E}[\theta|\theta > \theta_0]} < 0$, we let $E_2^*(\theta_0, \theta_1) = \mathbb{E}[\theta|\theta > \theta_0]$. Otherwise, we let $E_2^* \equiv E_2^*(\theta_0, \theta_1)$ be the solution to

$$V(E_2^*) - V(\mathbb{E}[\theta|\theta_0 < \theta < \theta_1]) = V'(E_2^*)(E_2^* - \mathbb{E}[\theta|\theta_0 < \theta < \theta_1]).$$

Next, consider the program:

$$\max_{\theta_0 \in [0, \overline{\theta}_0],\ \theta_1 \in [\overline{\theta}_0, 1]} \widehat{V}(\theta_0, \theta_1, E_2^*(\theta_0, \theta_1)) \text{ s.t. } \theta_1 \geq \theta_0.$$

Let θ_0^* and θ_1^* be the maximisers. There are three cases to consider:

(i) $\mathbb{E}[\theta|\theta > \theta_0^*] < E_2^*(\theta_0^*, \theta_1^*) < \mathbb{E}[\theta|\theta \geq \theta_1^*]$.

(ii) $E_2^*(\theta_0^*, \theta_1^*) = \mathbb{E}[\theta|\theta \geq \theta_1^*]$ because $\frac{\mathrm{d}\widehat{V}(\theta_0^*, \theta_1^*, \mathbb{E}[\theta|\theta \geq \theta_1^*])}{\mathrm{d}E_2} > 0$.

(iii) $E_2^*(\theta_0^*, \theta_1^*) = \mathbb{E}[\theta|\theta > \theta_0^*]$ because $\frac{\mathrm{d}\widehat{V}(\theta_0^*, \theta_1^*, \mathbb{E}[\theta|\theta \geq \theta_1^*])}{\mathrm{d}E_2} < 0$.

We consider case (i) first. Letting $E_2^* \equiv E_2^*(\theta_0, \theta_1)$ and $\alpha_{\theta_0, \theta_1}^* \equiv \alpha_{\theta_0, \theta_1, E_2^*(\theta_0, \theta_1)}$, the

Envelope theorem gives us that

$$\begin{aligned}
&\frac{\mathrm{d}\widehat{V}\left(\theta_1,\theta_1,E_2^*\right)}{\mathrm{d}\theta_1}=\left.\frac{\partial\widehat{V}\left(\theta_0,\theta_1,E_2\right)}{\partial\theta_1}\right|_{E_2=E_2^*}\\
&=f\left(\theta_1\right)\left(\theta_1-\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)\frac{1-F\left(\theta_0\right)}{F\left(\theta_1\right)-F\left(\theta_0\right)}\alpha^*_{\theta_0,\theta_1}\\
&\qquad\times\left[V'\left(\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)-\frac{V\left(E_2^*\right)-V\left(\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)}{E_2^*-\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]}\right]
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\mathrm{d}\widehat{V}\left(\theta_0,\theta_1,E_2^*\right)}{\mathrm{d}\theta_0}=\left.\frac{\partial\widehat{V}\left(\theta_0,\theta_1,E_2\right)}{\partial\theta_0}\right|_{E_2=E_2^*}\\
&=\frac{f\left(\theta_0\right)}{f\left(\theta_1\right)}\frac{\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]-\theta_0}{\theta_1-\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]}\frac{\mathrm{d}\widehat{V}\left(\theta_0,\theta_1,E_2^*\left(\theta_0,\theta_1\right)\right)}{\mathrm{d}\theta_1}\\
&\qquad+f\left(\theta_0\right)\left(\mathbb{E}\left[\theta|\theta>\theta_0\right]-\theta_0\right)\frac{V\left(E_2^*\right)-V\left(\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)}{E_2^*-\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]}\\
&\qquad-f\left(\theta_0\right)\left(\mathbb{E}\left[\theta|\theta>\theta_0\right]-\theta_0\right)\frac{V\left(\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)\alpha^*_{\theta_0,\theta_1}+V\left(E_2^*\right)\left(1-\alpha^*_{\theta_0,\theta_1}\right)}{\mathbb{E}\left[\theta|\theta>\theta_0\right]-\theta_0}.
\end{aligned}$$

Thus, at any interior optimal $\theta_0^*\in(0,\theta_0)$ and $\theta_1^*\in(\theta_0^*,1)$, it must be that

$$V'\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)\left(E_2^*-\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)=V\left(E_2^*\right)-V\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right),$$

which ensures that the impact on the average profit from workers equals the gains from more workers being assigned to $j=1$ net of losses from fewer workers being assigned to $j=2$; and

$$\frac{V\left(E_2^*\right)-V\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)}{E_2^*-\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]}=\frac{V\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)\alpha^*_{\theta_0^*,\theta_1^*}+V\left(E_2^*\right)\left(1-\alpha^*_{\theta_0^*,\theta_1^*}\right)}{\mathbb{E}\left[\theta|\theta>\theta_0^*\right]-\theta_0^*},$$

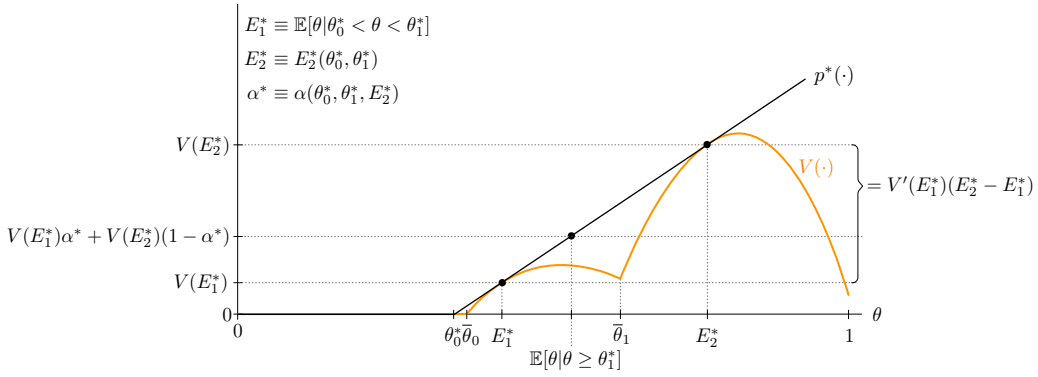
which ensures that changes in profits due to changes in workers assigned to $j=1$ or 2 equals the changes in profits from retaining fewer workers. If an interior solution $(\theta_0^*,\theta_1^*,E_2^*)$ to the first-order condition exists, then we can construct a convex supporting function p^* as

follows:

$$p^*(\theta) := \left[V'\left(\mathbb{E}\left[\theta|\theta_0^* < \theta < \theta_1^*\right]\right)(\theta - \theta_0^*)\right]_+$$
$$= \begin{cases} V(\theta) = 0 & \text{if } \theta \in [0, \theta_0^*], \\ V'\left(\mathbb{E}\left[\theta|\theta_0^* < \theta < \theta_1^*\right]\right)(\theta - \theta_0^*) & \text{if } \theta \in (\theta_0^*, 1]. \end{cases}$$

By construction, p^* is convex and $p^* \geq V$. One can also verify that p^* and $G_{\theta_0^*,\theta_1^*,E_2(\theta_0^*,\theta_1^*)}$ satisfy (18)–(20). Hence, we conclude that $G_{\theta_0^*,\theta_1^*,E_2^*(\theta_0^*,\theta_1^*)}$ solves (3). Observe that this corresponds to the case in which the interval $(\theta_0^*, 1)$ is bi-pooled to $\mathbb{E}[\theta|\theta_0^* < \theta < \theta_1^*]$ and $E_2(\theta_0^*, \theta_1^*)$. Figure 6 depicts the typical example of this case.

Figure 6: Case 1: $\mathbb{E}[\theta|\theta \geq \theta_0^*] < E_2^* < \mathbb{E}[\theta|\theta \geq \theta_1^*]$.



We now consider case (ii) in which $E_2^*(\theta_0^*, \theta_1^*) = \mathbb{E}[\theta|\theta \geq \theta_1^*]$. The expected payoff for the firm now is:

$$\widehat{V}\left(\theta_0^*, \theta_1^*, \mathbb{E}\left[\theta|\theta \geq \theta_1^*\right]\right) = V\left(\mathbb{E}\left[\theta|\theta_0^* < \theta < \theta_1^*\right]\right)\left(F\left(\theta_1^*\right) - F\left(\theta_0^*\right)\right)$$
$$+ V\left(\mathbb{E}\left[\theta|\theta \geq \theta_1^*\right]\right)\left(1 - F\left(\theta_1^*\right)\right)$$

Observe that

$$\frac{\mathrm{d}\widehat{V}\left(\theta_0, \theta_1, \mathbb{E}\left[\theta|\theta \geq \theta_1\right]\right)}{\mathrm{d}\theta_0}$$
$$= f(\theta_0)\left[\mathbb{E}\left[\theta|\theta_0 < \theta < \theta_1\right] - \theta_0\right]\left(V'\left(\mathbb{E}\left[\theta|\theta_0 < \theta < \theta_1\right]\right) - \frac{V\left(\mathbb{E}\left[\theta|\theta_0 < \theta < \theta_1\right]\right)}{\mathbb{E}\left[\theta|\theta_0 < \theta < \theta_1\right] - \theta_0}\right),$$

and

$$\begin{aligned}
&\frac{\mathrm{d}\widehat{V}\left(\theta_0,\theta_1,\mathbb{E}\left[\theta|\theta\geq\theta_1\right]\right)}{\mathrm{d}\theta_1}\\
&=f\left(\theta_1\right)\left(V\left(\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)+V'\left(\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)\left(\theta_1-\mathbb{E}\left[\theta|\theta_0<\theta<\theta_1\right]\right)\right)\\
&\quad+f\left(\theta_1\right)V'\left(\mathbb{E}\left[\theta|\theta\geq\theta_1\right]\right)\left(\mathbb{E}\left[\theta|\theta\geq\theta_1\right]-\theta_1\right)\\
&\quad-f\left(\theta_1\right)\left[V\left(\mathbb{E}\left[\theta|\theta\geq\theta_1\right]\right)\right].
\end{aligned}$$

Suppose $\theta_0^*>0$, then the first-order condition with respect to θ_0 means

$$V'\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]-\theta_0^*\right)=V\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right),$$

but if $\theta_0^*=0$, then

$$\frac{\mathrm{d}\widehat{V}\left(0,\theta_1^*,\mathbb{E}\left[\theta|\theta\geq\theta_1^*\right]\right)}{\mathrm{d}\theta_0}=V'\left(\mathbb{E}\left[\theta|\theta<\theta_1^*\right]\right)-\frac{V\left(\mathbb{E}\left[\theta|\theta<\theta_1^*\right]\right)}{\mathbb{E}\left[\theta|\theta<\theta_1^*\right]}<0.$$

The first-order condition with respect to θ_1 is

$$\begin{aligned}
V\left(\mathbb{E}\left[\theta|\theta\geq\theta_1^*\right]\right)=V&\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)\\
&+V'\left(\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)\left(\theta_1^*-\mathbb{E}\left[\theta|\theta_0^*<\theta<\theta_1^*\right]\right)\\
&+V'\left(\mathbb{E}\left[\theta|\theta\geq\theta_1^*\right]\right)\left(\mathbb{E}\left[\theta|\theta\geq\theta_1^*\right]-\theta_1^*\right)
\end{aligned}$$

Letting $E_1^*\equiv\mathbb{E}[\theta|\theta_0^*<\theta<\theta_1^*]$ and $E_2^*\equiv\mathbb{E}[\theta|\theta\geq\theta_1^*]$, construct p^* as follows:

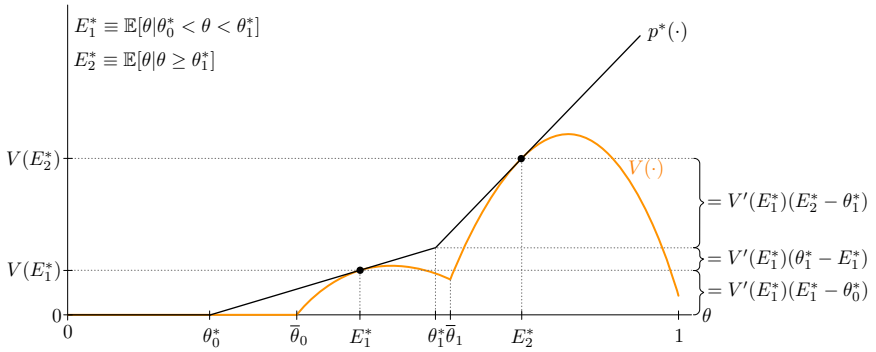
$$\begin{aligned}
p^*\left(\theta\right)&:=\max\left\{\left[V\left(E_1^*\right)+V'\left(E_1^*\right)\left(\theta-E_1^*\right)\right]_+,V\left(E_2^*\right)+V'\left(E_2^*\right)\left(\theta-E_2^*\right)\right\} \qquad (27)\\
&=\begin{cases}\left[V\left(E_1^*\right)+V'\left(E_1^*\right)\left(\theta-E_1^*\right)\right]_+ & \text{if } \theta\in[0,\theta_1^*],\\ V\left(E_2^*\right)+V'\left(E_2^*\right)\left(\theta-E_2^*\right) & \text{if } \theta\in(\theta_1^*,1].\end{cases}
\end{aligned}$$

In case $\theta_0^*>0$, the first-order condition with respect to θ_0 means that $V'(E_1^*)>0$ so that p^* is convex on $[0,\theta_1^*)$. If $\theta_0^*=0$, then p^* is affine (and thus convex) on $[0,\theta_1^*)$. To ensure that p^* is convex on $[0,1]$, we further require that $V'(E_2^*)\geq V'(E_1^*)$. By construction, $p^*\geq V$. Moreover, p^* and $G_{\theta_0^*,\theta_1^*,E_2^*(\theta_0^*,\theta_1^*)}$ satisfy (18)–(20) and so we conclude that $G_{\theta_0^*,\theta_1^*,E_2^*(\theta_0^*,\theta_1^*)}$ solves (3). Observe that, in this case, the interval (θ_0^*,θ_1^*) is pooled to E_1^*) and the interval $[\theta_1^*,1]$ is pooled to E_2^*. Figure 7 shows the typical example of the case when $E_2^*(\theta_0^*,\theta_1^*)=$

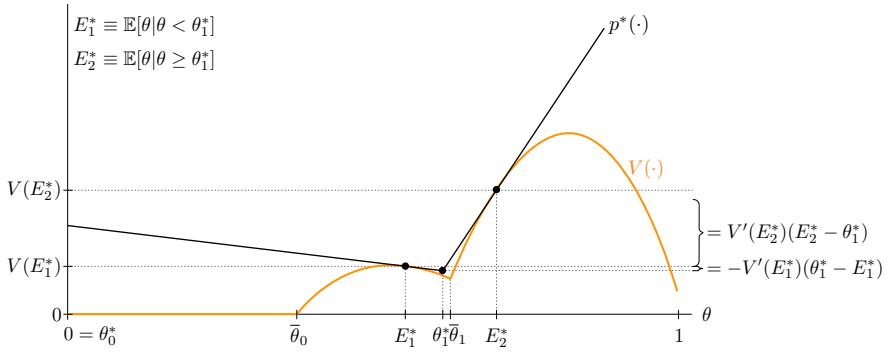
$\mathbb{E}[\theta|\theta \geq \theta_1^*]$ and either $\theta_0^* \in (0, \overline{\theta}_0)$ (Figure 7a) or $\theta_0^* = 0$ (Figure 7b).

Figure 7: Case 2: $E_2^*(\theta_0^*, \theta_1^*) = \mathbb{E}[\theta|\theta \geq \theta_1^*]$.

(a) Case 2a: $\theta_0^* \in (0, \overline{\theta}_0)$.



(b) Case 2b: $\theta_0^* = 0$.



Finally, we consider case (iii) in which $E_2^*(\theta_0^*, \theta_1^*) = \mathbb{E}[\theta|\theta > \theta_0^*]$. The expected payoff for the firm is now:

$$\widehat{V}\left(\theta_0^*, \theta_1^*, \mathbb{E}\left[\theta|\theta > \theta_0^*\right]\right) = V\left(\mathbb{E}\left[\theta|\theta > \theta_0^*\right]\right)\left(1 - F\left(\theta_0^*\right)\right).$$

Since the payoff does not depend on θ_1, we set it to $\theta_1^* = 1$. Observe that

$$\frac{\mathrm{d}\widehat{V}\left(\theta_0, \theta_1, \mathbb{E}\left[\theta|\theta > \theta_0\right]\right)}{\mathrm{d}\theta_0} = f\left(\theta_0\right)\left[\mathbb{E}\left[\theta|\theta > \theta_0\right] - \theta_0\right]\left(V'\left(\mathbb{E}\left[\theta|\theta > \theta_0\right]\right) - \frac{V\left(\mathbb{E}\left[\theta|\theta > \theta_0\right]\right)}{\mathbb{E}\left[\theta|\theta > \theta_0\right] - \theta_0}\right).$$

Thus, the first-order condition then requires that, if $\theta_0^* \in (0, \theta_0^*)$ and $\theta_1^* \in (\theta_0^*, 1)$,

$$V'\left(\mathbb{E}\left[\theta|\theta > \theta_0^*\right]\right)\left(\mathbb{E}\left[\theta|\theta > \theta_0^*\right] - \theta_0^*\right) = \frac{V\left(\mathbb{E}\left[\theta|\theta > \theta_0^*\right]\right)}{\mathbb{E}\left[\theta|\theta > \theta_0^*\right] - \theta_0^*};$$

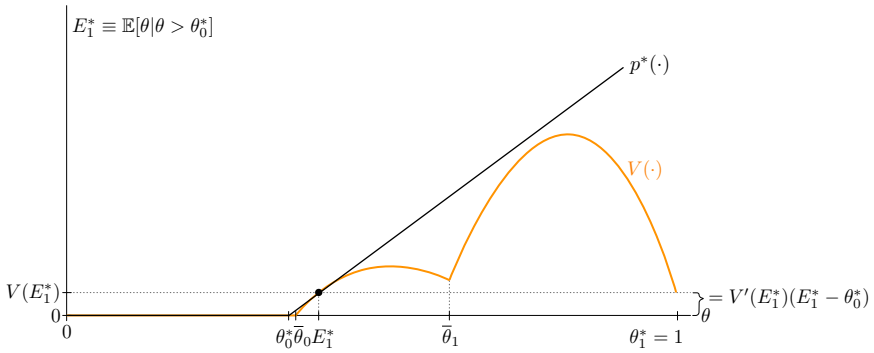
i.e., the tangent to V at $\mathbb{E}[\theta|\theta > \theta_0^*]$ crosses the x-axis at θ_0^*. Construct p^* now as follows:

$$p^*(\theta) = \begin{cases} V(\theta) = 0 & \text{if } \theta \in [0, \theta_0^*], \\ V'\left(\mathbb{E}\left[\theta|\theta \geq \theta_0^*\right]\right)\left(\theta - \theta_0^*\right) & \text{if } \theta \in (\theta_0^*, 1]. \end{cases}.$$

Observe that p^* is convex, $p^* \geq V$. Moreover, p^* and $G_{\theta_0^*, \theta_1^*, E_2^*(\theta_0^*, \theta_1^*)}$ satisfy (18)—(20) so that $G_{\theta_0^*, \theta_1^*, E_2^*(\theta_0^*, \theta_1^*)}$ solves (3). In this case, the interval $(\theta_0^*, 1]$ is pooled to $\mathbb{E}[\theta|\theta > \theta_0^*]$.

Figure 8: Case 3: $E_2^*(\theta_0^*, \theta_1^*) = \mathbb{E}[\theta|\theta > \theta_0^*]$.

(a) Case 3a: $E_1^* \in (\overline{\theta}_0, \overline{\theta}_1)$.



(b) Case 3b: $E_1^* \in (\overline{\theta}_1, 1)$.

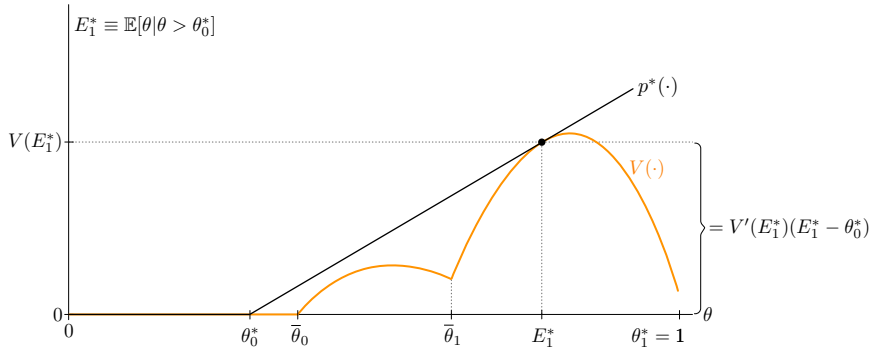
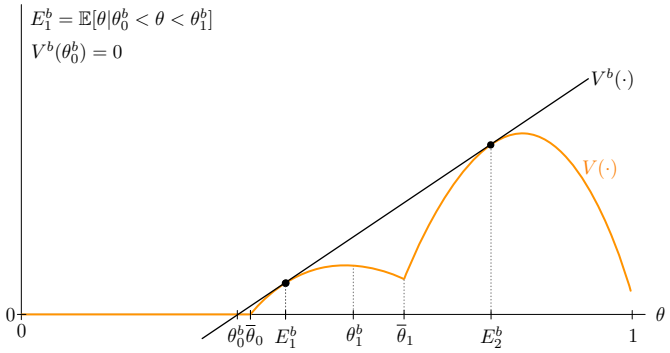


Figure 8 shows the typical examples of the case when $E_2^*(\theta_0^*, \theta_1^*) = \mathbb{E}[\theta|\theta > \theta_0^*]$. ■

We now characterise feasible G that solves the firm's information design problem, (4). Towards this goal, in case V has a bi-tangent, we denote it as $V^b(\cdot)$, and let E_1^b and E_2^b denote the two tangency points, where $E_2^b > E_1^b$, and let θ_0^b and θ_1^b be points such that $V^b(\theta_0^b) = 0$ and $\mathbb{E}[\theta|\theta_0^b < \theta < \theta_1^b] = E_1^b$. Figure 9 shows how E_1^b, E_2^b, θ_0^b and θ_1^b are obtained.

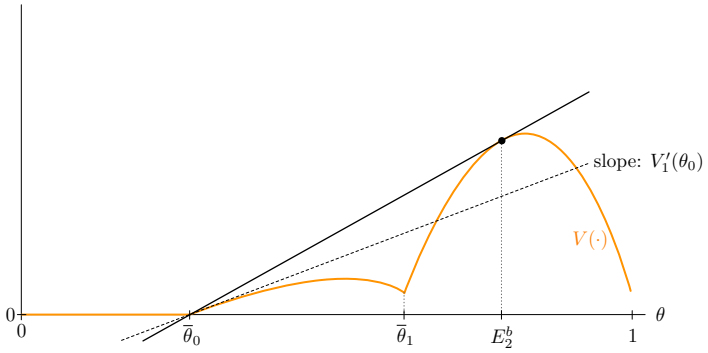
Figure 9: V has a bi-tangent, V^b.



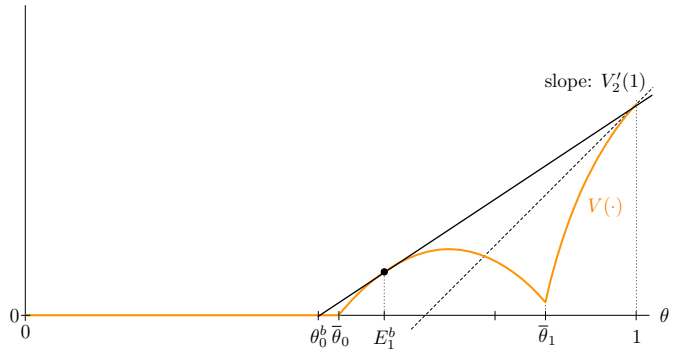
In case V does not have a bi-tangent, then at least one of (22) and (23) does not hold. If (22) does not hold, then let $E_2^b \in [\overline{\theta}_1, 1]$ be the point such that the tangent to V_2 at E_2^b goes through the point $(\overline{\theta}_0, 0)$. If (23) does not hold, then let $E_1^b \in [\overline{\theta}_0, \overline{\theta}_1]$ be the point such that the tangent to V_1 at E_1^b goes through the point $(1, V_2(1))$, and let θ_0^b be the point such that the tangents equals zero.

Figure 10: V does not have a bi-tagent.

(a) Case: (22) does not hold.



(b) Case: (23) does not hold



Theorem 3. *Suppose $p_2^o \geq p_1^o$ and $\mathbb{E}[\theta] < E_2^o$. Suppose a bi-tangent of V exists.*

(i) If

$$E_1^b \leq \mathbb{E}[\theta|\theta > \theta_0^b] \leq E_2^b \leq \mathbb{E}[\theta|\theta \geq \theta_1^b], \tag{28}$$

then an optimal G exists that bi-pools the interval $[\theta_0^b, 1]$ to E_1^b and E_2^b.

(ii) If (28) does not hold and

$$E_2^b > \mathbb{E}[\theta|\theta \geq \theta_1^b], \tag{29}$$

then an optimal G exists that pools the interval (θ_0^*, θ_1^*) to $\mathbb{E}[\theta|\theta_0^* < \theta < \theta_1^*] > E_1^b$, and pools the interval $[\theta_1^*, 1]$ to $E[\theta|\theta \geq \theta_1^*] < E_2^b$, where $\theta_0^* \in [0, \theta_0^b)$ and $\theta_1^* \in (\theta_0^*, 1)$.

(iii) If (28) does not hold and

$$E_1^b > \mathbb{E}[\theta|\theta > \theta_0^b], \tag{30}$$

then an optimal G exists that pools the interval $(\theta_0^*, 1]$ to $\mathbb{E}[\theta|\theta > \theta_0^*]$.

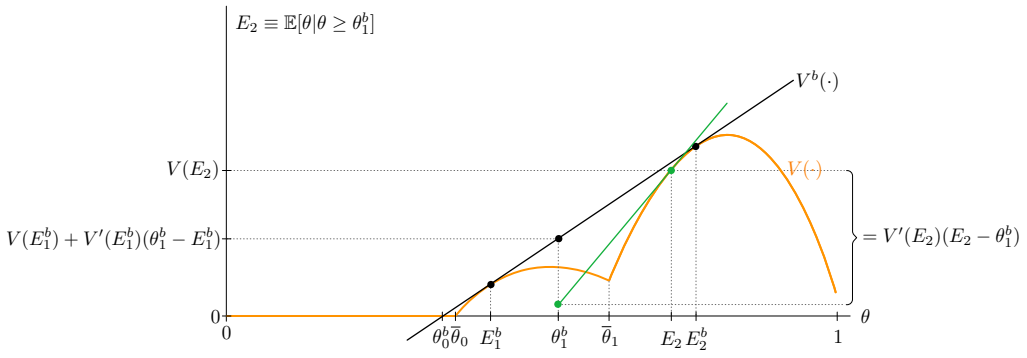
Suppose V does not have a bi-tangent: If (22) *does not hold, then optimal G has the same features as case (iii) above if $\mathbb{E}[\theta|\theta > \overline{\theta}_0] \geq E_2^b$ and otherwise G has the same features as case (ii).* (23) *does not hold, then optimal G has the same features as case (iii) above if $\theta_0^b < \mathbb{E}[\theta|\theta > \theta_0^b]$ and otherwise G has the same features as case (ii).*

Proof. (i) By Fact 2(ii), the interval $[\theta_0^b, 1]$ can be bi-pooled at E_1^b and E_2^b if and only if (28) holds. This case corresponds to case 1 from Lemma 6 with $\theta_0^* = \theta_0^b$, $\theta_1^* = \theta_1^b$ and $E_2^* = E_2^b$. (ii) This case includes the case in which (28) does not hold and $E_2^b > \mathbb{E}[\theta|\theta > \theta_0^b]$ because $\theta_1^b \geq \theta_0^b$. Consider first the case $E_2 \geq \overline{\theta}_1$. Letting denote $E_2 \equiv \mathbb{E}[\theta|\theta > \theta_0^b]$, we first argue that

$$V(E_2) < V(E_1^b) + V'(E_1^b)(\theta_1^b - E_1^b) + V'(E_2)(E_2 - \theta_1^b). \tag{31}$$

Toward a contradiction, suppose the inequality does not hold. Then, there must be a line through the point $(\theta_1^b, V(E_1^b) + V'(E_1^b)(\theta_1^b - E_1^b))$ that is tangent to V at $\mathbb{E}[\theta|\theta \geq \theta_1^b]$. However, by the choice of θ_1^b, V has a unique tangent that goes through the point $(\theta_1^b, V(E_1^b) + V'(E_1^b)(\theta_1^b - E_1^b))$ given by the tangent at E_2^b—a contradiction. Figure 11 depicts a typical example of this case.

Figure 11: Case (ii): Positive difference.



For any $\theta_0 \in [0, \theta_0^b]$, let $E_1(\theta_0)$ be given by

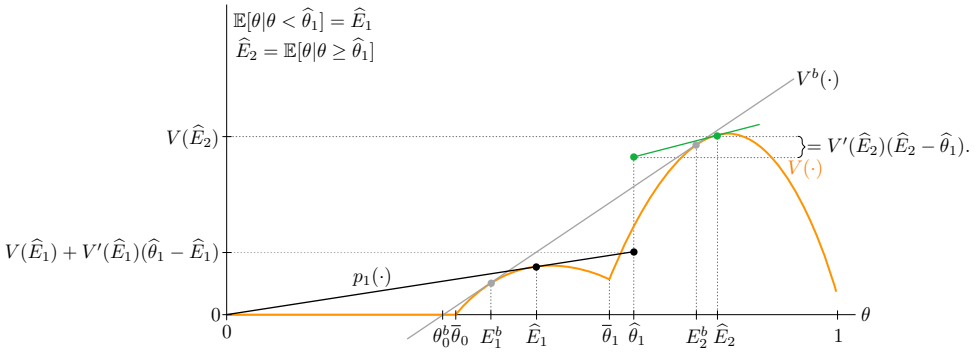
$$\theta_0 + V'(E_1(\theta_0))E_1(\theta_0) = V(E_1(\theta_0));$$

i.e., $E_1(\theta_0)$ is the point such that the line through $(\theta_0, 0)$ and $(E_1(\theta_0), V(E_1(\theta_0)))$ is tangent to V at $E_1(\theta_0)$. Because $E_1^b \in (\overline{\theta}_0, \overline{\theta}_1)$, concavity of V on $[\overline{\theta}_0, \overline{\theta}_1]$ means that $E_1(\theta_0)$ is continuous and strictly decreasing. Moreover, $\theta_1(\theta_0)$, defined implicitly by $\mathbb{E}[\theta|\theta_0 < \theta <$

$\theta_1(\theta_0)] = E_1(\theta_0)$, is also continuous and strictly decreasing. Now, let $p_1 : \theta \to \mathbb{R}$ denote the smallest linear function that exceeds V_1 and let $\widehat{E}_1$ denote the smallest $E \in \theta$ such that $p_1(E) = V_1(E)$. Let $\widehat{\theta}_1$ be such that $\mathbb{E}[\theta|\theta < \widehat{\theta}_1] = \widehat{E}_1$. Suppose, as shown in Figure 12,

$$V(\mathbb{E}[\theta|\theta \geq \widehat{\theta}_1]) > V(\widehat{E}_1) + V'(\widehat{E}_1)(\widehat{\theta}_1 - \widehat{E}_1) + V'(\mathbb{E}[\theta|\theta \geq \widehat{\theta}_1])(\mathbb{E}[\theta|\theta \geq \widehat{\theta}_1] - \widehat{\theta}_1). \quad (32)$$

Figure 12: Case (ii.a): Negative difference.



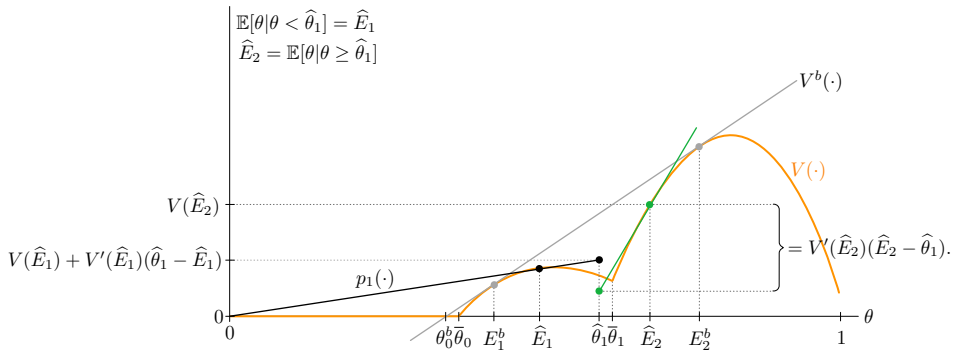
In this case, by the intermediate value theorem, there exists a $\theta_0 \in (0, \overline{\theta}_0)$ such that

$$\begin{aligned} V\left(\mathbb{E}\left[\theta|\theta \geq \theta_1\left(\theta_0\right)\right]\right) = V\left(E_1\left(\theta_0\right)\right) + V'\left(E_1\left(\theta_0\right)\right)\left(\theta_1\left(\theta_0\right) - E_1\left(\theta_0\right)\right) \\ + V'\left(\mathbb{E}\left[\theta|\theta \geq \theta_1\left(\theta_0\right)\right]\right)\left(\mathbb{E}\left[\theta|\theta \geq \theta_1\left(\theta_0\right)\right] - \theta_1\left(\theta_0\right)\right). \end{aligned}$$

Moreover, observe that such a θ_0 is unique due to the concavity of V on $[\overline{\theta}_0, \overline{\theta}_1]$ and on $[\overline{\theta}_1, 1]$. This case corresponds to the case 2a in the proof of Lemma 6.

Suppose now (32) does not hold as shown in Figure 13.

Figure 13: Case (ii.b): Positive difference.



Now let $\widetilde{\theta}_1 > \widehat{\theta}_1$ be such that $p_1(\widetilde{\theta}_1|\widetilde{\theta}_1) = V(\widetilde{\theta}_1)$, where

$$p_1\left(\theta'|\theta_1\right) \coloneqq V\left(\mathbb{E}\left[\theta|\theta<\theta_1\right]\right) + V'\left(\mathbb{E}\left[\theta|\theta<\theta_1\right]\right)\left(\theta' - \mathbb{E}\left[\theta|\theta<\theta_1\right]\right).$$

We will argue that (see Figure 14):

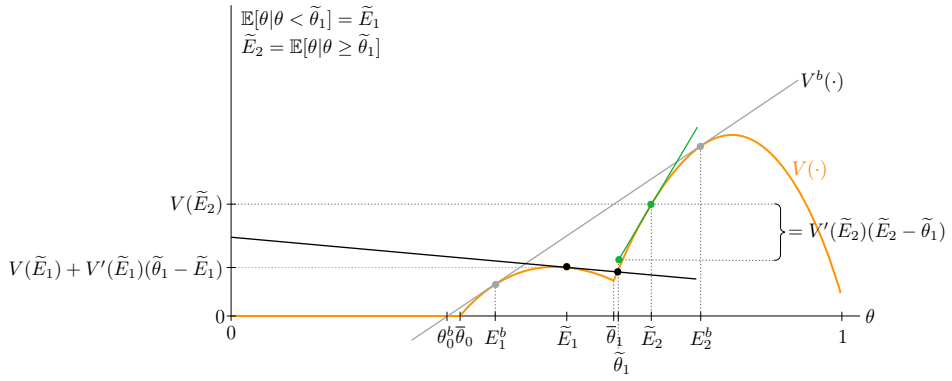
$$\begin{aligned} V(\mathbb{E}[\theta|\theta\geq\widetilde{\theta}_1]) > V(\mathbb{E}[\theta|\theta<\widetilde{\theta}_1]) + V'(\mathbb{E}[\theta|\theta<\widetilde{\theta}_1])(\widetilde{\theta}_1 - \mathbb{E}[\theta|\theta<\widetilde{\theta}_1]) \\ + V'(\mathbb{E}[\theta|\theta\geq\widetilde{\theta}_1])(\mathbb{E}[\theta|\theta\geq\widetilde{\theta}_1] - \widetilde{\theta}_1) \end{aligned}$$

By construction, $\widetilde{\theta}_1 \in (\overline{\theta}_1, 1)$ is such that the tangent to V at $\mathbb{E}[\theta|\theta<\widetilde{\theta}_1] \in (\overline{\theta}_1, 1)$ intersects V at $\widetilde{\theta}_1$. Since V is concave on $[\overline{\theta}_1, 1]$, $V'(\cdot)$ is strictly decreasing on $(\widetilde{\theta}_1, 1]$, for any $E_2 \in (\widetilde{\theta}_1, E_2^b) \subseteq (\widetilde{\theta}_1, 1]$,

$$V'(E_2)(E_2 - \widetilde{\theta}_1) > V(E_2) - V(\tilde{\theta}_1).$$

Letting $E_2 = \mathbb{E}[\theta|\theta<\widetilde{\theta}_1]$ gives the desired inequality.

Figure 14: Case (ii.b): Negative difference.



By the intermediate value theorem, there exists $\theta_1 \in (\widehat{\theta}_1, \widetilde{\theta}_1)$ such that

$$\begin{aligned} V\left(\mathbb{E}\left[\theta|\theta\geq\theta_1\right]\right) = V\left(\mathbb{E}\left[\theta|\theta<\theta_1\right]\right) + V'\left(\mathbb{E}\left[\theta|\theta<\theta_1\right]\right)\left(\theta_1 - \mathbb{E}\left[\theta|\theta<\theta_1\right]\right) \\ + V'\left(\mathbb{E}\left[\theta|\theta\geq\theta_1\right]\right)\left(\mathbb{E}\left[\theta|\theta\geq\theta_1\right] - \theta_1\right). \end{aligned}$$

Moreover, observe that such a θ_0 is unique due to the concavity of V on $[\overline{\theta}_0, \overline{\theta}_1]$ and on $[\overline{\theta}_1, 1]$. This case corresponds to case 2b in the proof of Lemma 6.

Now suppose that $\mathbb{E}[\theta|\theta\geq\theta_1^b] < \overline{\theta}_1$. Because V is concave on $[\overline{\theta}_0, \overline{\theta}_1]$, $\mathbb{E}[\theta|\theta\geq\theta_1^b] > E_1^b$ means that $V'(\mathbb{E}[\theta|\theta\geq\theta_1^b]) < V'(E_1^b)$. Hence, a function constructed as in (27) would not be convex. As already argued, choosing any $\theta_0 \in (0, \theta_0^b)$ means $E_1(\theta_0) > E_1^b$ and

$\theta_1(\theta_0) > \theta_1^b$ and $E_1(\theta_0) \geq \mathbb{E}[\theta|\theta \geq \theta_1^b]$. So long as $E_1(\theta_0) < \overline{\theta}_1$, a function constructed as in (27) would not be convex. If $E_1(\theta_0) \geq \overline{\theta}_1$ then we can proceed as above with θ_0 replacing $\overline{\theta}_0$ so we are once again in either case 2a or 2b of Lemma 6.

(iii) Let $\widetilde{\theta}_0$ be such that $\mathbb{E}[\theta|\theta > \widetilde{\theta}_0] = E_1^b$ and note that $\widetilde{\theta}_0 > \overline{\theta}_0$ because of condition (30). Define $\gamma : [\theta_0^b, \widetilde{\theta}_0] \to \mathbb{R}$ as

$$\gamma(\theta_0) := V'(\mathbb{E}[\theta|\theta > \theta_0]) - \frac{V(\mathbb{E}[\theta|\theta > \theta_0])}{\mathbb{E}[\theta|\theta > \theta_0] - \theta_0}.$$

Since the line segment connecting θ_0^b and E_1^b is tangent to V at E_1^b, $\widetilde{\theta}_0 > \theta_0^b$, and V is concave on $[\overline{\theta}_0, \overline{\theta}_1]$, the condition (30) implies that

$$\gamma(\theta_0^b) > 0 > \gamma(\widetilde{\theta}_0).$$

Since γ is continuous, by the intermediate value theorem, there exists $\theta_0 \in (\theta_0^b, \widetilde{\theta}_0)$ such that $\gamma(\theta_0^*) = 0$. Observe that this case corresponds to case (iii) in the proof of Lemma 6.

Now, suppose that a bi-tangent does not exist. Consider the first the case in which (22) does not hold. If $\mathbb{E}[\theta|\theta > \overline{\theta}_0] \geq E_2^b$. Since $\mathbb{E}[\theta] \leq \overline{E}$, there exists $\theta_0 \in [0, \overline{\theta}_0]$ such that

$$V\left(\mathbb{E}\left[\theta|\theta > \theta_0\right]\right) = V'\left(\mathbb{E}\left[\theta|\theta > \theta_0\right]\right)\left(\mathbb{E}\left[\theta|\theta > \theta_0\right] - \theta_0\right),$$

which is case (iii) in the proof of Lemma 6. Now suppose $\mathbb{E}[\theta|\theta > \overline{\theta}_0] < \overline{E}_0$. Observe that neither case (iii) nor case (i) in the proof of Lemma 6 is possible; i.e., only case (ii) can arise.

Finally, suppose (23) does not hold. If $\mathbb{E}[\theta|\theta > \theta_0^b] < E_1^b$, then any tangency to V on $[\overline{\theta}_0, 1]$ cannot be on $[\overline{\theta}_1, 1]$ so that this case corresponds to case (iii) in the proof of Lemma 6. If $\mathbb{E}[\theta|\theta > \theta_0^b] \geq E_1^b$, neither case (iii) nor case (i) in the proof of Lemma 6 is possible; i.e., only case (ii) can arise. ■

We now characterise the optimal G when $p_1^o \geq p_2^o$.

Theorem 4. *Suppose $p_1^o \geq p_2^o$. Then, V has a bi-tangent on $[\overline{\theta}_0, 1]$. Let E_1^b and E_2^b denote the two bi-tangency points.*

(i) If $\mathbb{E}[\theta] \in [0, E_1^o]$, then G associated with $(\theta_0, 1, 1)$ for some $\theta_0 \in [0, \overline{\theta}_0)$ solves the firm's problem. Moreover, no worker is ever assigned position $j = 2$.

(ii) *If* $\mathbb{E}[\theta] \in [E_1^o, E_1^b]$, *then* $G = F$ *solves the firm's problem and assigning all workers to position* $j = 1$ *is optimal.*

(iii) *If* $\mathbb{E}[\theta] \in [E_1^b, E_2^b]$...

(a) ... *and* (28) *holds with* $\theta_0^b = 0$,[45] *then an optimal* G *bi-pools the* Θ *to* E_1^b *and* E_2^b.

(b) ... *but* (28) *does not hold, then an optimal* G *pools the interval* $(0, \theta_1^*)$ *to* $\mathbb{E}[\theta|\theta < \theta_1^*] > E_1^b$, *and pools the interval* $[\theta_1^*, 1]$ *to* $E[\theta|\theta \geq \theta_1^*] < E_2^b$ *for some* $\theta_1^* \in \Theta$.

(iv) *If* $\mathbb{E}[\theta] \in [E_2^b, 1]$, *then* $G = F$ *solves the firm's problem and assigning all workers to position* $j = 2$ *is optimal.*

Proof. The result is immediate from previous arguments while observing that the slope of p_1^o must be greater than the tangent so that $\theta_1^b = 0$. ■

A.6 Inefficiencies

A.6.1 Retention

Formally, we say that an HR policy $\varphi : \Theta \to \Delta(\{0,1,2\})$ exhibits:

- *over-retention* if $\int_0^{\overline{\theta}_0} \varphi(0|\theta) \mathrm{d}F(\theta) < F(\overline{\theta}_0)$;
- *under-retention* if $\int_{\overline{\theta}_0}^1 \varphi(0|\theta) \mathrm{d}F(\theta) > 0$;
- *efficient retention* if $\int_0^{\overline{\theta}_0} \varphi(0|\theta) \mathrm{d}F(\theta) = F(\overline{\theta}_0)$ and $\int_{\overline{\theta}_0}^1 \varphi(0|\theta) \mathrm{d}F(\theta) = 0$.

Note that these definitions allow for an HR policy to exhibit both over-retention and under-retention.

Toward proving Theorem 1, we first characterise the case in which efficient retention is possible.

Lemma 7. *An optimal HR policy exists with efficient retention if and only if the efficient retention threshold* $\overline{\theta}_0$ *satisfies* $V_2'(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0]) \geq V_1'(\overline{\theta}_0)$ *and* $V_2(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0]) = V_2'(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0])(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0] - \overline{\theta}_0)$.

[45] As before, θ_1^b is given via $\mathbb{E}[\theta|\theta < \theta_1^b] = E_1^b$

Proof. Let G be a solution to the firm's information design problem and suppose it can be induced by an HR policy that exhibits efficient retention. By Fact 1, the support function p of G equals zero on $[0, \overline{\theta}_0]$. Since $p \geq V$, p must have a kink with a right-derivative greater than $V_1'(\overline{\theta}_0)$. Since $\text{supp}(G) \subseteq [\overline{\theta}_0, 1]$, by Fact 2(ii), p must be tangent to V at some $E \in (\overline{\theta}_0, 1]$. Because V_1 is concave, it must be that $E \in (\overline{\theta}_1, 1]$. The concavity of V_2 means that such a point E, if it exists, must be unique. Finally, observe that condition $V_2(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0]) = V_2'(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0])(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0] - \overline{\theta}_0)$ ensures $E = \mathbb{E}[\theta|\theta \geq \overline{\theta}_0]$. That the converse also holds follows from Fact 1. ■

Let us now prove the key lemma that delivers 1.

Lemma 8. *No optimal HR policy exhibits under-retention.*

Proof. Let φ be an optimal HR policy that exhibits under-retention. Then, there exists a subset $S \subseteq [\overline{\theta}_0, 1]$ who are let go with strictly positive probability. In the associated information design game, consider the G that is induced by φ. Suppose we modify G by fully revealing the types of workers in S whenever they are to be let go. Since $V(\theta) > 0$ for all $\theta \in S$, the firm can assign them to positions that maximise their output to obtain a strictly positive payoff, rather than a payoff of zero from letting go of them. Moreover, the firm's profits from retained workers are unaffected by this modification of G. Thus, the firm's payoff must be strictly higher with this modification. By Lemma 2, the firm can attain at least this payoff from an HR policy, which contradicts the optimality of φ. ■

We now prove that any optimal HR policy (monotone or otherwise) exhibits over-retention.

Theorem 1. *An optimal HR policy exists with efficient retention if and only if the efficient retention threshold $\overline{\theta}_0$ satisfies*

$$V_2'(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0]) \geq V_1'(\overline{\theta}_0) \tag{33}$$

and

$$V_2(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0]) = V_2'(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0])(\mathbb{E}[\theta|\theta \geq \overline{\theta}_0] - \overline{\theta}_0). \tag{34}$$

Moreover, under-retention is never optimal. Therefore, if conditions (33) *and* (34) *do not hold, then optimal HR policy exhibits over-retention.*

Proof. The first two statements are simply Lemma 7 and 8, respectively. If conditions (33) and (34) do not hold, then no HR policy with efficient retention is optimal. That is, any optimal HR policy exhibits inefficient retention. It must also exhibit no under-retention by Lemma 8. By definitions of efficient retention and under-retention, it follows that any optimal HR policy must exhibit over-retention. ■

A.6.2 Promotion

We say that an HR policy $\varphi : \Theta \to \Delta(\{0,1,2\})$ exhibits:

- *over-promotion* if $\int_0^{\overline{\theta}_1} \varphi(2|\theta)\mathrm{d}F(\theta) > 0$;
- *under-promotion* if $\int_{\overline{\theta}_1}^1 \varphi(2|\theta)\mathrm{d}F(\theta) < 1 - F(\overline{\theta}_1)$;
- *efficient promotion* if $\int_0^{\overline{\theta}_1} \varphi(2|\theta)\mathrm{d}F(\theta) = 0$ and $\int_{\overline{\theta}_1}^1 \varphi(2|\theta)\mathrm{d}F(\theta) = 1 - F(\overline{\theta}_1)$

Note again that these definitions allow for an HR policy to exhibit both over-retention and under-retention. Recall that any feasible G can be induced by a monotone HR policy and that, if G is pooling (vs bi-pooling), then it can be induced by a deterministic monotone HR policy.

We first note that if the degenerate distribution of posterior means (i.e., F) is optimal, then there is over-promotion (resp. under-promotion) if the firm assigns all workers to $j = 2$ (resp. $j < 2$) under the prior belief. Toward proving Theorem 2, let us consider the case in which $p_2^o > p_1^o$.

Proposition 8. *Suppose $p_2^o > p_1^o$. Suppose no optimal monotone HR policy is deterministic. Then, optimal monotone HR policies can exhibit only over-promotion, only under-promotion, or both over- and under-promotion.*

Proof. If no optimal monotone HR policy is deterministic, then optimal G must bi-pool the interval $[\theta_0^b, 1]$ to $\mathbb{E}[\theta|\theta_0^b < \theta < \theta_1^b]$ and $E_2 \in (\mathbb{E}[\theta|\theta > \theta_0^b], \mathbb{E}[\theta|\theta \geq \theta_1^b])$. A monotone HR policy that induces G is one that: (i) assigns $j = 2$ with probability one to workers whose types are in $[\theta_1^b, 1]$, and (ii) assigns $j = 1$ with some probability β to workers whose types are in $[\theta_0^b, \theta_1^b]$ and assigns $j = 2$ with complementary probability. If $\overline{\theta}_1 \geq \theta_1^b$, then this monotone HR policy exhibits only over-promotion. However, if $\overline{\theta}_1 < \theta_1^b$, this monotone HR policy exhibits both over-promotion (because it sometimes assigns $j = 2$ to workers whose types are in $[\theta_0^b, \overline{\theta}_1]$) and under-promotion (because it sometimes assigns $j = 1$ to workers whose types are in $[\overline{\theta}_1, \theta_1^b]$). To show that monotone HR policy can exhibit only

under-promotion, let $\tilde{t}_1^b$ be such that $\mathbb{E}[\theta|\theta \geq \tilde{t}_1^b] = E_2^b$. If $\mathbb{E}[\theta|\theta_0^b < \theta < \tilde{\theta}_1^b] < E_1^b$, following the same argument as in the proof of Lemma 2 in Arieli et al. (2023), we can induce a bi-pooling distribution by a monotone HR policy that: (i) assigns $j = 1$ with probability one to workers whose types are in $[\theta_0^b, \tilde{\theta}_1^b]$, and (ii) assigns $j = 1$ with some probability β to workers whose types are in $[\tilde{\theta}_1^b, 1]$ and assigns $j = 2$ with complementary probability. In this case, if $\overline{\theta}_1 \leq \tilde{t}_0^b$, there is only under-promotion. ■

We now focus on the case in which an optimal HR policy is deterministic. To state the result formally, let $\overline{E}_2 := \mathbb{E}[\theta|\theta \geq \overline{\theta}_1]$. When $\overline{E}_2 < E_2^b$, define $\overline{E}_1^e$ be such that

$$V\left(\overline{E}_1^e\right) + V'\left(\overline{E}_1^e\right)\left(\overline{\theta}_1 - \overline{E}_1^e\right) = V\left(\overline{E}_2\right) + V'\left(\overline{E}_2\right)\left(\overline{\theta}_1 - \overline{E}_2\right). \tag{35}$$

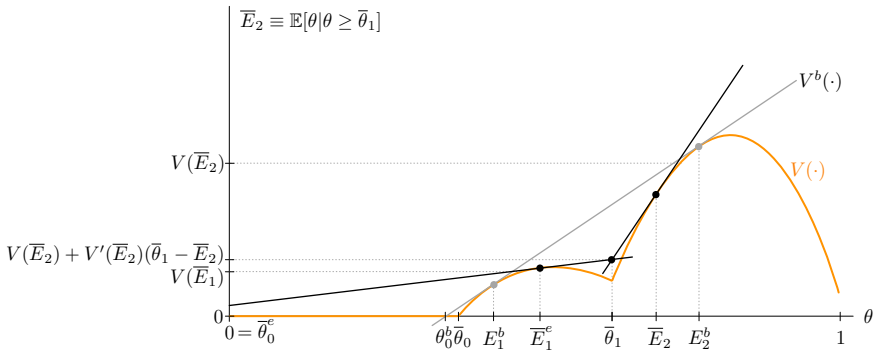
and let

$$\overline{\theta}_0^e = \begin{cases} 0 & \text{if } V\left(\overline{E}_1^e\right) \geq V'\left(\overline{E}_1^e\right)\overline{E}_1^e, \\ \overline{E}_1^e - \frac{V\left(\overline{E}_1^e\right)}{V'\left(\overline{E}_1^e\right)} & \text{otherwise.} \end{cases} \tag{36}$$

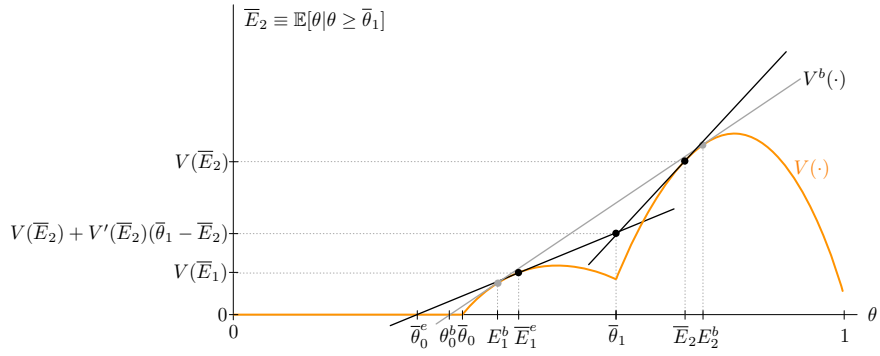
As Figure 15 shows, given $\overline{E}_2 < E_2^b$, $\overline{E}_1^e$ is the point such that the tangent to V at $\overline{E}_1^e$ intersects the tangent of V at $\overline{E}_2$ at $(\overline{\theta}_1, V(\overline{E}_2) + V'(\overline{E}_2)(\overline{\theta}_1 - \overline{E}_2))$. Figures 15a and 15b show the cases in which $\overline{\theta}_0^e = 0$ and $\overline{\theta}_0^e > 0$, respectively.

Figure 15: Over- vs under-promotion.

(a) Case: $\overline{\theta}_0^e = 0$.



(b) Case: $\overline{\theta}_0^e = \overline{E}_1^e - \frac{V(\overline{E}_1^e)}{V'(\overline{E}_1^e)}$.



Recall that we say that a deterministic monotone HR policy exhibits over-promotion (resp. under-promotion) if the associated promotion cutoff is below (resp. above) the efficient promotion threshold.

Proposition 9. *Suppose optimal monotone HR policy, φ^*, is deterministic.*

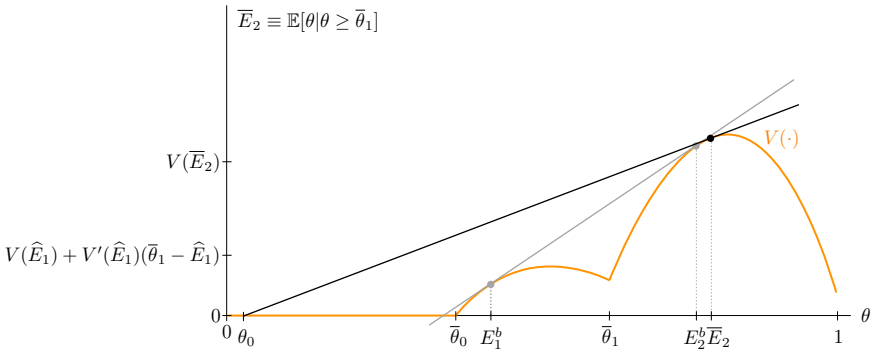
(i) Suppose V has a bi-tangent. Then, φ^* exhibits over-promotion if $E_2^b < \mathbb{E}[\theta|\theta \geq \theta_1^b]$ and either (a) $\overline{E}_2 > E_2^b$ or (b) $\overline{E}_2 < E_2^b$ and $\mathbb{E}[\theta|\overline{\theta}_0^e < \theta < \overline{\theta}_1] > \overline{E}_1^e$; otherwise, φ^* exhibits under-promotion.

(ii) Suppose no bi-tangent of V exists. Then, φ^* exhibits over-promotion if $\mathbb{E}[\theta|\overline{\theta}_0^e < \theta < \overline{\theta}_1] > \overline{E}_1^e$; otherwise φ^* exhibits under-promotion.

Proof. If V has a bi-tangent and a deterministic monotone HR policy is optimal, it must be the case that (28) does not hold and $E_2^b > \mathbb{E}[\theta|\theta \geq \theta_1^b]$.

Suppose $\overline{E}_2 > E_2^b$. Toward a contradiction, suppose $\theta_1^* \geq \overline{\theta}_1$. By definition of E_2^b as (one of) the bi-tangency points of V and the fact that $V|_{[\overline{\theta}_0, \overline{\theta}_1]}$ and $V|_{[\overline{\theta}_1, 1]}$ are concave, the tangent of V at $\overline{E}_2 \in (\overline{\theta}_1, 1)$ must lie strictly above V on $[\overline{\theta}_0, \overline{\theta}_1]$ and equal zero at some $\theta_0 < \overline{\theta}_0$ as can be seen from the figure below.

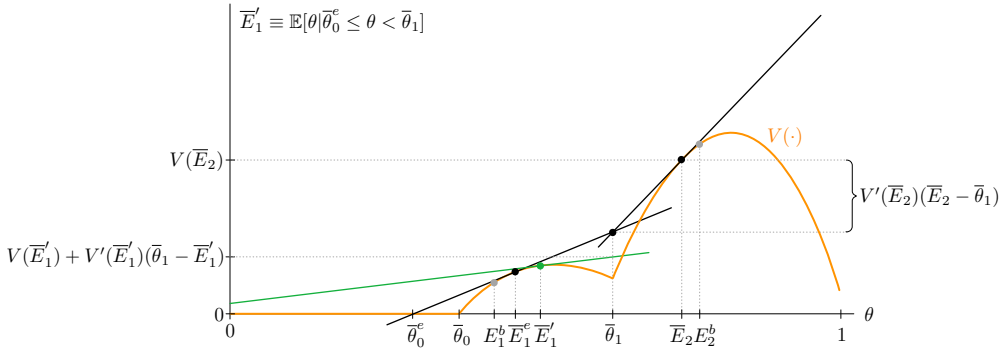
Figure 16: $\overline{E}_2 > E_2^b$: Over-promotion.



Thus, any G supported by p such that $p|_{[\overline{\theta}_1, 1]}$ is given by the tangent on V at $\overline{E}_2$ must pool the mass on $[\theta_0, 1]$ at $\mathbb{E}[\theta|\theta \geq \theta_0]$. But because $\mathbb{E}[\theta|\theta \geq \theta_0] < \overline{E}_2 \equiv \mathbb{E}[\theta|\theta \geq \overline{\theta}_1]$, such a support function is not a tangent to V at $\mathbb{E}[\theta|\theta \geq \theta_0]$; i.e., no such G is optimal. Hence, we must have $\theta_1^* < \overline{\theta}_1$.

Now suppose $\overline{E}_2 < E_2^b$. Consider first the case $\overline{E}_1' := \mathbb{E}[\theta|\overline{\theta}_0^e < \theta < \overline{\theta}_1] > \overline{E}_1^e$. Then, concavity of $V|_{[\overline{\theta}_0, \overline{\theta}_1]}$ implies that the tangent to V at $\overline{E}_1'$ evaluated at $\overline{\theta}_1$ must lie strictly below the intersection of the tangents to V at $\overline{E}_1^e$ and $\overline{E}_2$as can be seen in the figure below.

Figure 17: $\overline{E}_2 < E_2^b$ and $\overline{E}_1' > \overline{E}_1^e$: Over-promotion.

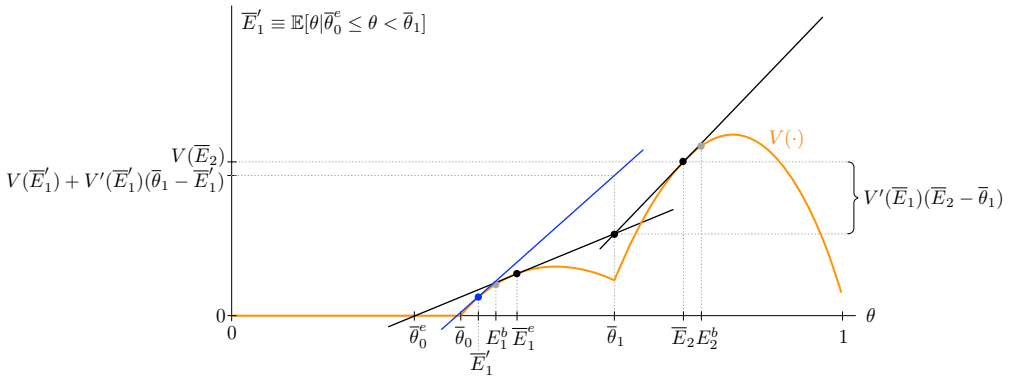


But this means that

$$\frac{\mathrm{d}\widehat{V}\left(\overline{\theta}_0^e, \overline{\theta}_1, \overline{E}_2\right)}{\mathrm{d}\theta_1}$$
$$= f\left(\overline{\theta}_1\right)\left[\left(V\left(\overline{E}_1'\right) + V'\left(\overline{E}_1'\right)\left(\overline{\theta}_1 - \overline{E}_1'\right)\right) - \left(V\left(\overline{E}_2\right) - V'\left(\overline{E}_2\right)\left(\overline{E}_2 - \overline{\theta}_1\right)\right)\right] \tag{37}$$

is strictly negative and so $\theta_1^* < \overline{\theta}_1$; i.e., there is over-promotion. If $\overline{E}_1' < \overline{E}_1^e$, analogous arguments yields that $\theta_1^* > \overline{\theta}_1$ (see figure below) so that there is under-promotion.

Figure 18: $\overline{E}_2 < E_2^b$ and $\overline{E}_1' < \overline{E}_1^e$: Under-promotion.



The case for when $E_2^b < \mathbb{E}[\theta|\theta \ge \theta_1^b]$ follows immediately from Theorem 3.

For the case in which V does not have a bi-tangent, observe that the argument for why there is over- or under-promotion when $\overline{E}_2 < E_2^b$ did not rely on bi-tangents. Hence, the same argument still goes through even when V does not have a bi-tangent. ■

We note that the condition $\overline{E}_2 > E_2^b$ holds if there is a sufficient mass of workers whose types are above E_2^b. The condition $\mathbb{E}[\theta|\overline{\theta}_0^e \le \theta < \overline{\theta}_1] > \overline{E}_1^e$ holds if there is a sufficient mass of workers whose types are in $[\overline{E}_1^e, \overline{\theta}_1]$.

Theorem 2. *An optimal deterministic monotone HR policy exhibits either over-promotion or under-promotion but not both. If no deterministic monotone HR policy is optimal, then an optimal monotone HR policy could exhibit only over-promotion, only under-promotion, or both.*

Proof. Case: $p_2^o > p_1^o$. Proposition Proposition 8 deals with the case when no optimal HR policy is optimal, and Proposition Proposition 9 deals with the case when optimal HR policy is deterministic.

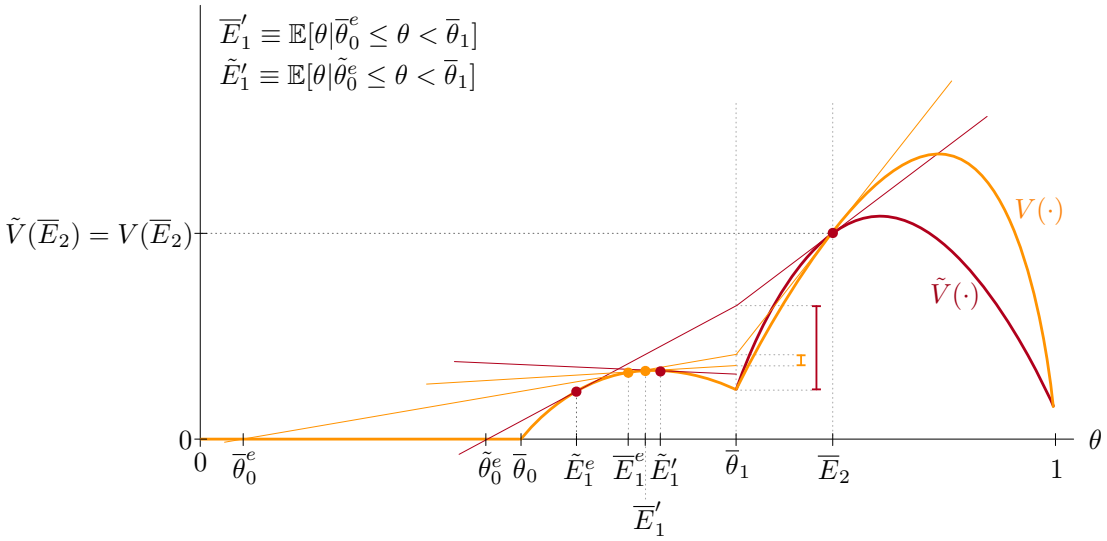
Case: $p_2^o \leq p_1^o$. From Theorem 4, the only interesting case when $p_1^o \geq p_2^o$ is if $\mathbb{E}[\theta] \in [E_1^b, E_2^b]$. In this case, the analysis of whether there is over- or under-promotion is the same as the case in which $p_2^o > p_1^o$ when the firm does not let go of any workers. ■

A.6.3 Comparative statics on c

Proposition 3. *Suppose no deterministic monotone HR policy is optimal. Consider comparable (v_2, c, F) and $(v_2, \tilde{c}, F)$ such that $c(\theta) = \tilde{c}(\theta)$ for $\theta \in [0, \overline{\theta}_1] \cup \{\overline{E}_2, 1\}$. Then, over-promotion is more likely under $(v_2, \tilde{c}, F)$ than under (v_2, c, F) if $\tilde{c}'(\overline{E}_2) > c'(\overline{E}_2)$.*

Proof. Suppose $\overline{E}_2 < E_2^b$. Fix comparable primitives (v_2, c, F) and $(v_2, \tilde{c}, F)$ that satisfy the hypothesis of the proposition and $\tilde{c}'(\overline{E}_2) > c'(\overline{E}_2)$. Let $\tilde{V}(\cdot) := [\max_{j \in \{1,2\}} v_j(\cdot) - \tilde{c}(\cdot)]_+$. Observe that V and $\tilde{V}$ coincide on $[0, \overline{\theta}_1] \cup \{\overline{E}_2, 1\}$. Moreover, $\tilde{V}'(\overline{E}_2) = k_2 - \tilde{c}'(\overline{E}_2) < k_2 - c'(\overline{E}_2) = V'(\overline{E}_2)$ and $\tilde{V}(\overline{E}_2) = V(\overline{E}_2)$ mean that the tangent to $\tilde{V}$ at $\overline{E}_2$ cuts $\overline{\theta}_1$ above the tangent to V at $\overline{E}_2$ (see figure below which shows the case when $\mathbb{E}[\theta | \overline{\theta}_0^e < \theta < \overline{\theta}_1] > \overline{\theta}_1^e$).

Figure 19: Comparative statics on c.



Given concavity of V on $[\overline{\theta}_0, \overline{\theta}_1]$, we must have $\tilde{E}_1^e < \overline{E}_1^e$, where $\tilde{E}_1^e$ solves (35) (with V replaced with $\tilde{V}$). In turn, $\tilde{\theta}_0^e$ defined via (36) (with $\overline{E}_1^e$ replaced with $\tilde{E}_1^e$) must be such that $\tilde{\theta}_0^e \geq \overline{\theta}_0^e$. Therefore, $\tilde{E}_1' := \mathbb{E}[\theta | \tilde{\theta}_0^e < \theta < \overline{\theta}_1] \geq \overline{E}_1' := \mathbb{E}[\theta | \overline{\theta}_0^e < \theta < \overline{\theta}_1]$. Hence, the derivative (37) must decrease. That is, over-promotion becomes more likely. ■

A.6.4 Comparative statics on v_2

Proposition 4. *Suppose no deterministic monotone HR policy is optimal. Consider comparable (v_2, c, F) and $(\tilde{v}_2, c, F)$. Then, over-promotion is no more likely under $(\tilde{v}_2, c, F)$ than under (v_2, c, F).*

Proof. Recall that $\overline{\theta}_1 \equiv \frac{s_1 - s_2}{k_2 - k_1}$ and consider an increase in k_2 to $\tilde{k}_2 = k_2 + \varepsilon$ for some $\varepsilon > 0$. Then, to ensure that the efficient promotion threshold remains unchanged, we need

$$\overline{\theta}_1 = \frac{s_1 - \tilde{s}_2}{\tilde{k}_2 - k_1} \Leftrightarrow \tilde{s}_2 = s_1 - \overline{\theta}_1 (k_2 - k_1) - \varepsilon \overline{\theta}_1 = s_2 - \varepsilon \overline{\theta}_1.$$

Let $\tilde{V}$ denote the firm's value function under $(\tilde{k}_2, \tilde{s}_2)$:

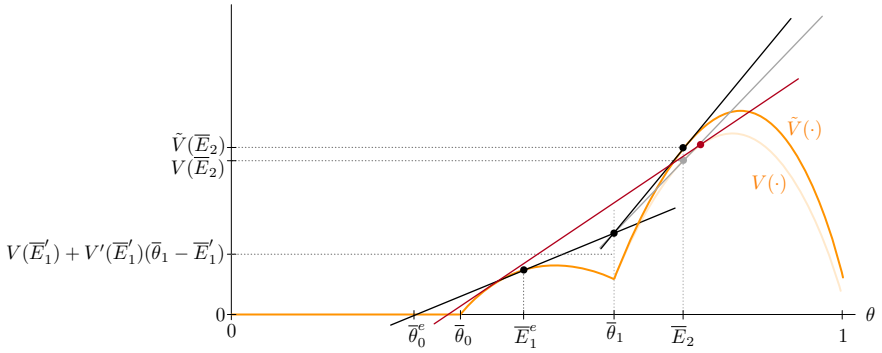
$$\tilde{V}(\theta) = \begin{cases} V(\theta) & \text{if } \theta \leq \overline{\theta}_1, \\ \tilde{s}_2 + \tilde{k}_2 \theta - c(\theta) = v_2(\theta) - c(\theta) + \varepsilon \left(\theta - \overline{\theta}_1\right) & \text{if } \theta > \overline{\theta}_1. \end{cases}$$

Note that the tangent to $\tilde{V}$ at $\overline{E}_2 \equiv \mathbb{E}[\theta | \theta \geq \overline{\theta}_1]$ is given by

$$\begin{aligned} \tilde{V}'\left(\overline{E}_2\right)\left(\theta - \overline{E}_2\right) + \tilde{V}\left(\overline{E}_2\right) &= \left(\tilde{k}_2 - c'\left(\overline{E}_2\right)\right)\left(\theta - \overline{E}_2\right) + \tilde{v}_2\left(\overline{E}_2\right) - c\left(\overline{E}_2\right) \\ &= \left(k_2 - c'\left(\overline{E}_2\right)\right)\left(\theta - \overline{E}_2\right) + \varepsilon\left(\theta - \overline{E}_2\right) \\ &\quad + v_2\left(\overline{E}_2\right) - c\left(\overline{E}_2\right) + \varepsilon\left(\overline{E}_2 - \overline{\theta}_1\right) \\ &= V'\left(\overline{E}_2\right)\left(\theta - \overline{E}_2\right) + V\left(\overline{E}_2\right) + \varepsilon\left(\theta - \overline{\theta}_1\right). \end{aligned}$$

Hence, the tangent to $\tilde{V}$ at $\overline{E}_2$ and the tangent to V at $\overline{E}_2$ intersect exactly at $\overline{\theta}_1$. But then the derivative (37) does not change. This is shown in the figure below.

Figure 20: Comparative statics on v_2.



Note further that since $\tilde{V}'|_{[\overline{\theta}_1,1]} > \tilde{V}|_{[\overline{\theta}_1,1]}$, the bi-tangency points of $\tilde{V}$ (if they exist), denoted $\tilde{E}_1^b$ and $\tilde{E}_2^b$, satisfy $\tilde{E}_1^b < E_1^b$ and $\tilde{E}_2^b > E_2^b$. Because $\overline{E}_2 < E_2^b$, it must be $\overline{E}_2 < \tilde{E}_2^b$ so that the nature of inefficiency remains unchanged. ■

A.6.5 Comparative statics on F

The following is a well-known result that tells us that if two CDFs F and H are ordered according to their likelihood ratios, then their conditional distributions are first-order stochastically ordered.[46]

Fact 3. *Let F and H be two continuous CDF with support $[0,1]$. Then,*

$$F(\cdot|a<X<b) \geq H(\cdot|a<X<b) \ \forall a,b \in [0,1] : b > a$$

if and only if $\frac{h(\cdot)}{f(\cdot)}$ is increasing.

It follows that if F and H are two prior distributions of workers' types and that $\frac{h(\cdot)}{f(\cdot)}$ is increasing (i.e., H dominates F in the likelihood ratio order), then $\mathbb{E}_H[\theta|\theta_0 < \theta < \theta_1] \geq \mathbb{E}_F[\theta|\theta_0 < \theta < \theta_1]$ for any $\theta_0, \theta_1 \in [0,1]$ with $\theta_1 > \theta_0$.

Proposition 5. *Suppose no deterministic monotone HR policy is optimal. Consider comparable (v_2, c, F) and $(v_2, c, \tilde{F})$. Then, over-promotion is more likely under $(v_2, c, \tilde{F})$ than under (v_2, c, F) if $\tilde{F}$ dominates F in the likelihood ratio order.*

Proof. By definition, that $\tilde{F}$ dominates F in the likelihood ratio order means $\frac{\tilde{f}(\cdot)}{f(\cdot)}$ is increasing. Hence, by Fact 3, $\overline{E}_2(\tilde{F}) := \mathbb{E}_{\tilde{F}}[\theta|\theta \geq \overline{\theta}_1] \geq \overline{E}_2(F) := E_F[\theta|\theta \geq \overline{\theta}_1]$. Moreover, because V is concave on $[\overline{\theta}_1, 1]$,

$$V\left(\overline{E}_2\left(\tilde{F}\right)\right) - V'\left(\overline{E}_2\left(\tilde{F}\right)\right)\left(\overline{E}_2\left(\tilde{F}\right) - \overline{\theta}_1\right) > V\left(\overline{E}_2(F)\right) - V'\left(\overline{E}_2(F)\right)\left(\overline{E}_2(F) - \overline{\theta}_1\right).$$

Define $\overline{\theta}_0^e(\tilde{F})$ and $\overline{\theta}_0^e(F)$ via (36) respectively for $\overline{E}_2(\tilde{F})$ and $\overline{E}_2(F)$. By concavity of V on $[\overline{\theta}_0, \overline{\theta}_1]$, we must have $\overline{\theta}_0^e(\tilde{F}) > \overline{\theta}_0^e(F)$. Hence,

$$\mathbb{E}_F[\theta|\overline{\theta}_0^e(F) < \theta < \overline{\theta}_1] \leq \mathbb{E}_F[\theta|\overline{\theta}_0^e(\tilde{F}) < \theta < \overline{\theta}_1] \leq \mathbb{E}_{\tilde{F}}[\theta|\overline{\theta}_0^e(\tilde{F}) < \theta < \overline{\theta}_1],$$

where the last inequality follows from Fact 3. Hence, the derivative (37) must be smaller under $\tilde{F}$ than under F. That is, over-promotion is more likely under $\tilde{F}$. ■

[46] See, for example, Shaked and Shanthikumar (2007), Theorem 1.C.5.

A.7 Partial commitment

Lemma 9. *Any HR policy that solves (2) must be deterministic and monotone.*

Proof. Fix an HR policy, φ. The result follows if we can prove the claim: there exists measurable subsets $S, S' \subseteq \theta$ such that (i) $F(S) = F(S') = \ell > 0$; (ii) $\sup S' \leq \widehat{\theta} := \inf S$; and (iii) $\varphi(\cdot|\theta) = \delta_{j'}$ for all $\theta \in S$ and $\varphi(\cdot|\theta') = \delta_j$ for all $\theta' \in S'$ with $j > j'$; then the firm's direct payoff is weakly greater from $\widehat{\varphi}$ that equals φ except that $\varphi(\cdot|\theta) = \delta_j$ for all $\theta \in S$ and $\varphi(\cdot|\theta') = \delta_{j'}$ for all $\theta' \in S'$. Condition (i) means that S and S' have the same size, (ii) means that S' consists of workers whose types are below that of any worker in S, and (iii) means that φ assigns a higher position to lower types of workers who are in S'. The claim says that the firm is better off by assigning lower types of workers to a lower position. To prove the claim, observe that $v_j(\theta)$ has increasing differences in (j, θ) since, for any $\theta, \theta' \in \theta$ such that $\theta' > \theta$, we have

$$v_2\left(\theta'\right) - v_2\left(\theta\right) = k_2\left(\theta' - \theta\right) > k_1\left(\theta' - \theta\right) = v_1\left(\theta'\right) - v_1\left(\theta\right) > 0 = v_0\left(\theta'\right) - v_0\left(\theta\right).$$

In particular, this means that, for any $j > j'$,

$$v_j\left(\theta\right) - v_{j'}\left(\theta\right) > v_j\left(\theta'\right) - v_{j'}\left(\theta'\right).$$

This observation, together with conditions (i) and (ii), means that

$$\begin{aligned}
\int_S \left[v_j\left(\theta\right) - v_{j'}\left(\theta\right)\right] \mathrm{d}F\left(\theta\right) &\geq \int_S \left[v_j\left(\widehat{\theta}\right) - v_{j'}\left(\widehat{\theta}\right)\right] \mathrm{d}F\left(\theta\right) \\
&= \left[v_j\left(\widehat{\theta}\right) - v_{j'}\left(\widehat{\theta}\right)\right] \ell \\
&= \int_{S'} \left[v_j\left(\widehat{\theta}\right) - v_{j'}\left(\widehat{\theta}\right)\right] \mathrm{d}F\left(\theta\right) \\
&\geq \int_{S'} \left[v_j\left(\theta\right) - v_{j'}\left(\theta\right)\right] \mathrm{d}F\left(\theta\right).
\end{aligned}$$

In fact, what we have shown is

$$\int_S v_j\left(\theta\right) \mathrm{d}F\left(\theta\right) + \int_{S'} v_{j'}\left(\theta\right) \mathrm{d}F\left(\theta\right) \geq \int_S v_{j'}\left(\theta\right) \mathrm{d}F\left(\theta\right) + \int_{S'} v_j\left(\theta\right) \mathrm{d}F\left(\theta\right);$$

i.e., the payoff from swapping assignments is greater. To see why the claim implies the result, observe that if φ is not monotone, then we can find $S, S' \subseteq \theta$ that satisfies the conditions of the claim. Now suppose that φ is monotone and involves randomisation, then there will

exist measures of workers who are sometimes assigned to a higher position and otherwise assigned to a lower position. However, the firm can improve its payoff by assigning the workers to a higher position without changing the marginal distribution via the claim. ■